

Experimentelles Design für Informationssuche im Cue-basierten probabilistischen Entscheiden

**Abhandlung zur Erlangung der Doktorwürde
der Philosophischen Fakultät
der Universität Zürich**

vorgelegt von Stephan Christen, MuttENZ /BL

**Angenommen im Herbstsemester 2008
auf Antrag von
Prof. Dr. Damian Läge und
Prof. Dr. Wolfgang Marx**

Zürich, 2009

Inhaltsverzeichnis

Zusammenfassung	1
Abstract	1
Vorwort und Danksagung	3
1 Einleitung	11
1.1 Das Informationboard	12
1.2 Die drei Building-Blocks: Such-, Stopp- und Entscheidungsregel	13
1.3 Ursprüngliche Experimental-Designs zur Untersuchung von TTB	14
1.3.1 Informationssuche in der Umwelt.....	14
1.3.2 Informationssuche im Gedächtnis.....	18
1.4 Weiterentwicklung des Informationboard anhand spezifischer Fragestellungen.....	21
1.4.1 Die Fragestellungen	21
1.4.2 Das Optionsschein-Szenario - Die Bedeutung aktiver Informationssuche	22
1.4.3 Das Quizshow-Szenario – Unterscheidung von Informationssuch-Strategien mit flexibel anbietbaren Validitäten und Diskriminationsraten	23
1.4.4 Das Äffchen-Szenario – Erlernbarkeit von Cue-Qualitäten	25
1.4.5 Das Trinkwasser-&-Börsen-Szenario – Zeitliche Stabilität von Cues & Zielvariablen / Zeit als Ressource.....	27
1.5 Literatur.....	28
Teil I - Optionsschein-Szenario	31
2 Ein Optionsscheinszenario zur Untersuchung der Strategien sequentieller Informationssuche: Aufbau, Ablauf und technische Spezifikation der Experimentalumgebung	33
2.1 Einleitung.....	34
2.2 Allgemeine Struktur des Experiments	35
2.3 Experimentalablauf.....	39
2.3.1 Überblick: Der Ablauf im Groben	39
2.3.2 Detaillierte Betrachtung eines Versuchsdurchgangs.....	44
2.4 Technische Spezifikationen	46
2.5 Diskussion.....	49
2.6 Literatur.....	50
3 Der Einfluss der Informationssuche auf die Entscheidung	51
3.1 Einleitung.....	52
3.2 Methoden	55

3.3	Resultate	61
3.3.1	Generelles Anlageverhalten: Bedingung mit vorgegebener Information.....	61
3.3.2	Generelles Suchverhalten: Bedingung mit aktiver Informationssuche	61
3.3.3	Übereinstimmung im Anlageverhalten zwischen und innerhalb der Versuchspersonen (Hauptfragestellung)	62
3.3.4	Vertrauen in Information (Fragestellung 2)	63
3.3.5	Interindividuelle Unterschiede im Suchverhalten (Fragestellung 3).....	65
3.3.6	Zusammenhang zwischen Anlageverhalten und Suchstrategien (Fragestellung 4)	65
3.4	Diskussion	67
3.5	Literatur.....	70

Teil II - Quizshow-Szenario 73

4	Die Quizshow: Ein experimentelles Design zur Messung von individuellen Informationssuch-Strategien in einfachen Urteilsheuristiken.....	75
4.1	Einleitung: Methodologische Überlegungen zur Messung des Entscheidungsverhaltens	76
4.2	Anforderungen an das Versuchsdesign	79
4.2.1	Bestimmung der angewendeten Informationssuch-Strategie	79
4.2.2	Aktive Informationssuche in 'realer Welt'	79
4.2.3	Standardisierung der realen Welt	80
4.3	Umsetzung der Anforderungen	80
4.3.1	Schwierige Wissensfragen	80
4.3.2	Telefonjoker	80
4.3.3	Vier Antwort-Alternativen	81
4.3.4	Natürliche Präsentationsform der entscheidungspsychologischen Konzepte	81
4.3.5	Feedback und Motivation der Versuchspersonen	81
4.3.6	Konfidenz-Rating	82
4.3.7	Messen der Suchregel in zwei Phasen.....	83
4.3.8	Kostenstruktur (Punkte vs. CHF-Beträge)	84
4.3.9	Ablauf des Experiments im Überblick	85
4.4	Der Versuchsablauf.....	85
4.4.1	Übersicht über die einzelnen Schritte des durchgeführten Quizshow-Experiments	85
4.4.2	Detaillierte Beschreibung der einzelnen Schritte des Experiments.....	86
4.5	Ablauf einer Aufgabe	95
4.6	Vorversuch zur Auswahl der Wissensfragen	103
4.6.1	Auswahl des Aufgaben-Basismaterials (Fragen und Antwort-Alternativen).....	103
4.6.2	Vorversuch	104
4.6.3	Antwortmuster-Analyse	105
4.6.4	Definitiver Fragepool	105

4.7	Erzeugung der Cues und Konstruktion der Cue-Sets.....	107
4.8	Mittlere Diskriminationsrate der Cues.....	111
4.9	Bestimmung der in Phase 2 zu testenden Suchstrategien	113
4.10	Performance-Abschätzung	115
4.11	Abschliessende Bemerkungen	118
4.12	Literatur.....	119
5	Was macht einen „guten Cue“ aus? Strategien der Informationssuche beim heuristischen Entscheiden unter Unsicherheit.....	121
5.1	Einleitung.....	122
5.1.1	Take The Best als ökologisch rationale Entscheidungsheuristik	122
5.1.2	Empirische Evidenz: Take The Best-Entscheidungen nur bei Informationskosten	124
5.1.3	Informationssuche und die Frage nach dem „guten Cue“	125
5.2	Experiment 1	128
5.2.1	Aufbau des Experiments	128
5.2.2	Resultate.....	130
5.2.3	Diskussion.....	132
5.3	Experiment 2	132
5.3.1	Aufbau des Experiments	133
5.3.2	Resultate.....	135
5.3.3	Diskussion.....	137
5.4	Gesamtdiskussion	138
5.5	Literatur.....	141
6	Und sie lernen es doch! Das Einschätzen der Validität von probabilistischen Hinweis-Cues und der Einfluss der subjektiven Repräsentation auf das Urteilsverhalten.....	145
6.1	Einleitung.....	146
6.2	Methoden	151
6.3	Resultate.....	155
6.3.1	Güte der gelernten Cues und ihrer Cue-Validitäten (Fragestellung 1).....	155
6.3.2	Kein Unterschied in den Repräsentationen zwischen den beiden Bedingungen	156
6.3.3	Bewertung der subjektiv repräsentierten Cue-Validitäten	158
6.3.4	Einfluss der wahrgenommenen Cue-Güte auf die Anwendung von TTB (Fragestellung 2)	158
6.3.5	Anwendung der Cue-Repräsentationen im Sinne von TTB.....	159
6.3.6	Individuelles Entscheidungsverhalten.....	159
6.4	Diskussion.....	161
6.4.1	Implikationen für die Plausibilität von TTB	164
6.5	Literatur.....	164

Teil III - Äffchen-Szenario	167
7 Das „Äffchen-Szenario“: Technische Spezifikation einer Experimentalumgebung zur Untersuchung der Lernbarkeit von Validität, Diskriminationsrate und Success von Informationsquellen	169
7.1 Einleitung	170
7.2 Aufbau des Experiments	173
7.3 Ablauf	173
7.4 Datengrundlage	198
7.5 Benutzeroberflächen	199
7.5.1 Login	199
7.5.2 Lern- und Anwendungsphase	199
7.5.3 Befragung	203
7.5.4 Nachbefragung	204
7.5.5 Persönliche Angaben	205
7.6 Details zur Programmierung	206
7.7 Auswertemöglichkeiten	207
7.8 Anhänge	209
7.8.1 Rahmengeschichte	209
7.8.2 Instruktionen der drei Expeditionen im Wortlaut	210
7.8.3 Fragestellungen der Befragungsphase im Wortlaut	211
7.8.4 Festgelegte Cue-Informationen	212
7.9 Literatur	214
8 Ist ein Gefühl für Informationsgüte lernbar? Ein Lernexperiment zu Validität, Diskriminationsrate und Success von Informationsquellen	215
8.1 Einleitung	216
8.2 Methode	225
8.2.1 Aufbau des Experimentes	225
8.3 Geplante Auswertung	231
8.3.1 Stichprobe	232
8.4 Resultate	232
8.4.1 Auswertung 1 und 2: Lernen des besten Cues (One-Sample-T-Test, einseitig) und Lernfähigkeit – Erfüllungsscores bei tatsächlichem Verhalten	232
8.4.2 Auswertung 3: Adaptivität des Suchverhaltens: Erfüllungsscores bei transferiertem Verhalten	235
8.4.3 Auswertung 4: Erfüllungsscores und Adaptivität des Suchverhaltens basierend auf der subjektiven Cue-Güte	237
8.4.4 Auswertung 5: Abschätzung des Lernens anhand deklarativer Angaben (Rangordnungen, Pearson-Korrelationen und Differenzbeträge)	238
8.4.5 Auswertung 6: Entscheidungsebene (Anzahl korrekter Entscheidungen)	240
8.5 Diskussion	240

8.6	Literatur.....	244
Teil IV - Trinkwasser- &-Börsen-Szenario		
9 Trinkwasser- und Börsenszenario: Aufbau, Ablauf und technische Spezifikation einer Experimentalumgebung zur entscheidungsvorbereitenden Informationssuche in zeitlichen Verläufen		245
9.1	Einleitung.....	248
9.2	Experimentalaufbau	249
9.2.1	Allgemeine Struktur des Experiments	249
9.2.2	Experimentalbedingungen des Trinkwasser-Szenarios im Detail.....	253
9.2.3	Experimentalbedingungen des Börsenszenarios im Detail	256
9.3	Experimentalablauf.....	259
9.3.1	Überblick über den Ablauf.....	259
9.3.2	Informationssuche (Bereich „Analyse“ bzw. „Informationen“)	262
9.3.3	Entscheidungsphase (Bereich „Handlung“ bzw. „Handel“)	263
9.3.4	Feedback (Bereich „Übersicht“)	264
9.4	Technische Spezifikation und Datenbasis.....	264
9.4.1	Technische Spezifikation	264
9.5	Datenbasis	268
9.6	Schlussbemerkungen.....	269
9.7	Literatur.....	270
10 Die zeitliche Dynamik von Umwelten: Heuristisches Wahlverhalten am Beispiel eines Trinkwasser-Szenarios mit und ohne Zeitachse.....		271
10.1	Einleitung.....	272
10.2	Methode	275
10.2.1	Aufbau des Experimentes	275
10.2.2	Operationalisierte Hypothesen.....	278
10.2.3	Versuchsdurchführung.....	280
10.3	Resultate.....	280
10.4	Diskussion.....	286
10.5	Schlussdiskussion	290
10.6	Literatur.....	292
11 Entscheiden in zeitlich dynamischen Umwelten: Suchen und Stoppen mit unterschiedlich stabilen Cues		295
11.1	Einleitung.....	296
11.1.1	Entscheiden im Zeitverlauf.....	296
11.1.2	Such- und Stoppregel in zeitlich dynamischen Umwelten	298
11.1.3	Hypothesen	299
11.2	Methode	299

11.2.1	Aufbau des Experiments	299
11.2.2	Versuchsdurchführung	302
11.2.3	Stichprobe	303
11.3	Resultate	303
11.3.1	Kaufhäufigkeiten aller Cues.....	303
11.3.2	Kaufhäufigkeiten einzelner Cues	303
11.3.3	Kaufhäufigkeiten auf individueller Ebene	306
11.3.4	Suchregel.....	307
11.3.5	Stoppverhalten.....	309
11.4	Diskussion	313
11.5	Literatur.....	316
	Curriculum Vitae	319

Zusammenfassung

Zur computergestützten Untersuchung verschiedener entscheidungspsychologischer Fragestellungen, in deren Zentrum jeweils die Informationssuche in Cue-basierten probabilistischen Entscheidungssituationen steht, haben wir vier neue Experimental-Umgebungen (Szenarien) entwickelt. Diese Szenarien werden in der vorliegenden Arbeit einerseits aus technischer, andererseits aus anwendungsorientierter Perspektive (d.h. anhand konkret damit durchgeführter Experimente) vorgestellt. Die vier Szenarien (Optionsschein-, Äffchen-, Quizshow-, Trinkwasser-&-Börsen-Szenario) sind auf die Untersuchung spezifischer Aspekte des Informationssuch-Verhaltens ausgerichtet. Sie basieren aus experimentalmethodischer Sicht auf einem Informationboard-Ansatz, der jeweils spezifisch für die zu klärenden Fragestellungen adaptiert werden musste. In allen Szenarien lösen die Versuchspersonen innerhalb eines spielerischen, performance-orientierten Kontexts eine Reihe gleichartiger Entscheidungsaufgaben: Aus einer Menge von Alternativen müssen sie jeweils die hinsichtlich eines vorgegebenen Kriteriums Beste auswählen. Da sie über die Alternativen nichts wissen, müssen sie zunächst nach Cues bzw. Hinweisen suchen, anhand derer sie die Qualität der Alternativen beurteilen können, bevor sie eine Entscheidung treffen. Der gewinnorientierte Kontext der Szenarien bewirkt eine Limitierung der Informationssuche, so dass anhand der aufgezeichneten Aktionen der Versuchsperson (insbesondere der Informationssuchreihenfolge) Rückschlüsse auf das Such- und Entscheidungsverhalten möglich sind.

Abstract

Four new computer-based experimental scenarios focusing on psychological processes during information search in cue-based, probabilistic decision situations were developed. For each scenario there is a chapter on its technical aspects and one or two on the conducted experiments based on this specific scenario. All four scenarios ("Optionsscheins-Szenario", "Äffchen-Szenario", "Quizshow-Szenario", "Trinkwasser-&-Börsen-Szenario") examine specific aspects of information search behaviour. From a technical point of view each scenario contains an information board adapted to the requirements of the research problems posed, where subjects accomplish a sequence of similar decision tasks within a playful, performance-oriented context: Out of several alternatives the one that matches a given criterion best has to be chosen. Initially no information on the alternatives is provided. Therefore the subjects first have to search for cues helping them to evaluate the quality of the alternatives before choosing one. The scenarios are embedded in profit-oriented contexts where every bit of information costs a certain amount of play money. This leads to a limitation of the amount of information searched for. Preventing subjects from consulting each and every available hint is a basic prerequisite for drawing valid conclusions (about search and decision strategies applied by the subjects) based on information search orders.

Vorwort und Danksagung

Im Alltag stehen wir häufig vor dem Problem, aus einer Menge von Alternativen die richtige bzw. die für unsere Zwecke am besten geeignete zu finden und auszuwählen. Oftmals liegen über die Alternativen zunächst nur einige wenige oder noch gar keine Informationen vor, auf die wir unsere Wahl abstützen könnten. Dementsprechend ist die Suche nach (weiteren) entscheidungsrelevanten Informationen ein wichtiger Bestandteil von Entscheidungsprozessen. Ein Beispiel: Anhand verschiedener Krankheitssymptome muss ein Arzt herausfinden, an welcher Krankheit sein Patient leidet. Einige Symptome, wie etwa geschwollene Hände oder rote Pusteln, sind für den Arzt rasch ersichtlich, währenddem er andere, wie beispielsweise eine erhöhte Körpertemperatur, erst noch in Erfahrung bringen muss. Aufgrund der gesammelten Informationen wird der Arzt schliesslich eine Annahme darüber treffen, von welcher Krankheit sein Patient betroffen ist.

Zur Beschaffung einer Information müssen wir entweder Zeit oder Geld (oder beides zusammen) aufwenden. Beide Ressourcen stehen uns in der Regel nur in begrenztem Umfang zur Verfügung. Dementsprechend beenden wir die Informationssuche sobald wir keine weiteren Ressourcen mehr für die Informationssuche aufwenden können oder wollen (beispielsweise weil wir mit der Informationslage zufrieden sind oder den antizipierten potentiellen Nutzen zusätzlicher Informationen als zu gering erachten). Da der Arzt in unserem Beispiel ausreichend Zeit hat, beendet er die Suche nach Informationen darüber, ob bestimmte Symptome vorliegen oder nicht, sobald er sich hinreichend sicher fühlt um eine akkurate Diagnose treffen zu können.

In der Terminologie der Entscheidungspsychologie kann die oben geschilderte Situation wie folgt beschrieben werden: Ein Entscheider (Arzt) sucht für die Entscheidungsfindung (Diagnosestellung) nach verschiedenen probabilistischen Cues¹ (Krankheitssymptomen) und schliesst anhand der gefundenen Cue-Informationen (Informationen darüber, ob bestimmte Symptome beim Patienten vorhanden sind oder nicht) darauf, dass Objekt XY (Krankheitsdiagnose XY) unter allen möglichen Objekten (Krankheitsdiagnosen) hinsichtlich der Zielvariable (Diagnosegüte) den höchsten Ausprägungsgrad bzw. Kriteriumswert besitzt (die bestmögliche Krankheitsdiagnose darstellt).

Unsere Entscheidungsexperimente laufen üblicherweise auf einem Computer ab, auf dessen Bildschirm ein „Informationboard“ präsentiert wird. Dieses enthält einerseits die zur Auswahl stehenden Objekte und Cues und dient andererseits der Anzeige der objektbezogenen Informationen (Cue-Informationen), auf deren Grundlage sich die Versuchspersonen schliesslich für eines der Objekte entscheiden sollen. Diese Informationen müssen entweder aktiv gesucht werden

¹ Im klassischen Sinn ist ein Cue ein Hinweisreiz, der durch seine Präsenz oder Abwesenheit auf das Bestehen einer besonderen Situation hinweist und dadurch zu einem entscheidungs- bzw. handlungs-relevanten Faktor werden kann – sofern er von der handelnden Person wahrgenommen wird.

oder können bereits von Beginn weg offen vorliegen. Sofern die Informationen aktiv gesucht werden müssen, enthält das Board anstelle der Informationen zunächst nur die Cues. Diese können als Prädiktorvariablen (oder als probabilistische Informationsquellen) betrachtet werden, die, wenn sie von den Versuchspersonen konsultiert (abgerufen / befragt) werden, wichtige Informationen über eines oder mehrere der zur Auswahl stehenden Objekte liefern, auf die sie sich bei ihrer Entscheidung abstützen können. Anhand der Ausprägungen der Prädiktorvariablen kann in der Regel nicht mit Sicherheit auf die Ausprägungen der Zielvariablen geschlossen werden, weshalb die Informationsquellen bzw. Cues als probabilistisch bezeichnet werden. Von einem Cue-Hinweis sprechen wir dann, wenn die konkrete Ausprägung einer Prädiktorvariablen bekannt ist. Die Wahrscheinlichkeit (oder eben Probabilität) mit der anhand eines Cue-Hinweises eine korrekte Entscheidung getroffen werden kann, wird als Validität (V) dieses Cues bezeichnet und kann (sofern bekannt) in der Form eines Prozentwerts angegeben werden. Die von einem Cue gelieferten Hinweise sind nicht immer brauchbar. Wenn ein Cue nicht diskriminiert d.h. wenn er für alle zur Auswahl stehenden Objekte identische Hinweise liefert, trägt dies nichts zur Unterscheidung zwischen den Entscheidungsalternativen bei. Die Häufigkeit, mit der ein Cue diskriminiert, kann (sofern bekannt) ebenfalls in der Form eines Prozentwerts angegeben werden und wird als Diskriminationsrate (D) bezeichnet. Wenn von einem Cuegüte-Mass die Rede ist, meint man damit entweder V oder D, oder ein aus V und D abgeleitetes Mass zur Beschreibung des Cue-Nutzens (Cuegüte). Die Cuegüte ist also eine vom verwendeten Gütemass abhängige Zahl, die angibt, wie häufig man im Mittel eine korrekte Entscheidung treffen wird, wenn man sich dabei einzig auf diesen Cue abstützt. Die beiden bekanntesten Masse für die Cuegüte sind (neben V und D) die Usefulness ($D \cdot V$) und der Success ($SUC = D \cdot V + [1 - D] \cdot \text{guess}$). Die Usefulness eines Cues gibt an, wie häufig man im Mittel von diesem Cue einen nützlichen Hinweis erhält, d.h. einen Hinweis, der zu einer korrekten Entscheidung führt – sofern man sich hinweiskonform entscheidet. Verglichen mit der Usefulness berücksichtigt Success zusätzlich die Möglichkeit, durch Raten zu einer korrekten Entscheidung zu gelangen (wobei guess die Ratewahrscheinlichkeit² bezeichnet). Dies betrifft allerdings nur diejenigen Fälle, in denen der konsultierte Cue nicht diskriminiert³ (1-D). In vielen Experimenten gibt man den Versuchspersonen die genauen Validitäten der Cues (und ggf. auch deren Diskriminationsraten) bekannt. Es besteht aber auch die Möglichkeit, die Versuchspersonen diese Werte in einer dem eigentlichen Experiment vorge-schalteten Trainingsphase lernen zu lassen.

Da die Informationssuche in den von uns durchgeführten entscheidungspsychologischen Experimenten jeweils von zentraler Bedeutung ist, führt das Computerprogramm im Hintergrund jeweils Buch über den Prozess der Informationssuche (d.h. darüber, welche Cues angeklickt werden, in welcher Reihenfolge dies geschieht und welche Entscheidung schliesslich getroffen wird). Diese Informationen sind für die Auswertungen im Hinblick auf die Fragestellungen i.d.R. von zentraler Bedeutung. Um zu verhindern, dass die Versuchspersonen in jeder Aufgabe ein-

² Die Ratewahrscheinlichkeit ist abhängig von der Anzahl zur Auswahl stehender Entscheidungsalternativen. Im Fall von n Alternativen ist $\text{guess} = 1/n$ also 50% im Zweialternativenfall, 25% im Vieralternativenfall, etc.

³ Sofern ein Cue diskriminiert, sollte man sich grundsätzlich hinweiskonform entscheiden – d.h. für die durch den Hinweis angezeigte Alternative.

fach immer sämtliche zur Verfügung stehenden Cues konsultieren, erfolgt die Informationssuche häufig unter Aufwendung einer monetären Ressource (Spielgeld), und die Aufgabe der Versuchspersonen (über das ganze Experiment gesehen) besteht in der Maximierung der Anzahl korrekt getroffener Entscheidungen bei gleichzeitiger Minimierung des Ressourcenaufwands.

Eine Aufgabe in einem Entscheidungsexperiment könnte zum Beispiel so aussehen, dass im Informationboard die Namen zweier historischer Persönlichkeiten – bspw. „Ulrich Zwingli“ und „Albert Einstein“ (= Menge der Objekte) – angezeigt werden und diejenige Person ausgewählt werden soll, die das höhere Lebensalter erreichte (= Kriterium bzw. Zielvariable). Besitzt eine Versuchsperson kein geeignetes Vorwissen über diese Personen, um die Aufgabe sofort lösen zu können, so muss sie nach Informationen⁴ suchen, um in dieser Situation mehr tun zu können als einfach nur zu raten. Mit anderen Worten: Sie muss einen oder mehrere Cues konsultieren, die ihr für diesen Zweck im Experiment zur Verfügung gestellt werden. In unserem Beispiel könnten dies die folgenden drei Cues sein: 1) Tod durch Unfall (Ja/Nein); 2) Gewaltsamer Tod (Ja/Nein); 3) Nach 1900 geboren (Ja/Nein)⁵. Eine „Ja“-Antwort spricht dabei jeweils eher für eine kürzere Lebensdauer und eine „Nein“-Antwort für eine längere. Würde eine Versuchsperson alle drei Cues konsultieren, so erhielte sie lediglich von Cue 2 (Gewaltsamer Tod) einen brauchbaren Hinweis, da einzig dieser Cue für die beiden Personen unterschiedliche Aussagen macht (Ulrich Zwingli: Ja; Albert Einstein: Nein). Die beiden anderen Cues diskriminieren nicht: Sie liefern jeweils für beide Personen ein „Nein“. Aufgrund dieser Hinweise entscheidet sich die Versuchsperson für Einstein – und liegt damit richtig⁶.

In klassischen Entscheidungsexperimenten wurden solche wahrscheinlichkeitsbehafteten Hinweise den Versuchspersonen oft einfach vorgegeben, ohne dass sie zuvor aktiv nach diesen Informationen hätten suchen müssen. Die Aufgabe der Versuchspersonen bestand in diesen Experimenten in erster Linie darin, die vorliegenden Informationen so zu integrieren, dass sie möglichst viele gute (korrekte) Entscheidungen damit erzielten. Dadurch dass man sämtliche (relevanten) Informationen bereits vorgab, schloss man den Prozess der aktiven Informationssuche von der experimentellen Untersuchung aus.⁷ Möglicherweise verhalten sich Personen aber

⁴ Bei diesen Informationen handelt es sich natürlich nicht um konkrete Angaben zum erreichten Lebensalter, sondern lediglich um indirekte Hinweise, die einen relativen Grössenvergleich (hinsichtlich des erreichten Lebensalters) zulassen oder bestenfalls auf die ungefähren, absoluten Geburts- und Todeszeitpunkte schliessen lassen.

⁵ Das 19. Jahrhundert brachte enorme Fortschritte in der Diagnose und Therapie vieler Krankheiten. Vor allem durch die Entwicklungen im Bereich der Naturwissenschaften (Histologie, Bakteriologie, Mikrobiologie, etc.). Innerhalb von wenigen Jahrzehnten konnten die Erreger vieler vorher kaum erfolgreich behandelbarer Krankheiten gefunden werden (1826: Diphtherie; 1849: Milzbrand; 1873: Lepra; 1882: Tuberkulose; 1894: Pest, etc.).

⁶ Albert Einstein (1897–1955), wurde mit seinen 76 Jahren 29 Jahre älter als Ulrich Zwingli (1484–1531), der im Zweiten Kappelerkrieg katholischen Innerschweizern in die Hände fiel und von diesen ermordet wurde.

⁷ Liegen sämtliche Informationen bereits von Anfang an offen vor, sind zwar keine Aussagen über das Informationssuchverhalten mehr möglich, anhand der getroffenen Entscheidungen können aber immerhin noch Schlussfolgerungen zum Informations-Integrationsverhalten gezogen werden. Voraussetzung dafür ist allerdings, dass die für das Experiment benötigten Ausgangsdaten (d.h. die Cue-Struktur, die von den Cues abgegebenen Hinweise, etc.) vorgängig geeignet festgelegt worden sind.

systematisch anders, wenn sie ihre Entscheidungen auf der Basis vorgegebener Informationen treffen, als wenn sie – bei identischer Informationsstruktur im Hintergrund – aktiv nach diesen Informationen suchen müssen. Aus diesem Grund entwickelten wir ein Experimental-Design, das Optionsschein-Szenario, mit dem wir die Bedeutung der aktiven Informationssuche für die spätere Entscheidung untersuchen können. Das Optionsschein-Szenario wird in Teil 1 dieser Arbeit ausführlich vorgestellt. Es lässt sowohl die attributweise als auch die alternativenweise Suche nach Informationen zu und bietet dadurch die Möglichkeit zur vertieften Untersuchung interpersoneller Unterschiede und individueller Präferenzen bei der aktiven Informationssuche.

Wenn ein Experiment (wie im obigen Beispiel) auf realen Daten bzw. Umwelten basiert, läuft man Gefahr, dass die Versuchspersonen die richtigen Antworten teilweise bereits von vornherein kennen (d.h. Vorwissen besitzen) oder zu kennen glauben und deshalb nur spärlich nach Informationen suchen. Dies wiederum hat zur Folge, dass wegen der geringen Datenmenge, die dabei anfällt, nicht mehr nachvollziehbar ist, aufgrund welcher Informationen die Versuchspersonen ihre Entscheidungen getroffen haben. Dadurch sind keine Rückschlüsse auf die individuell angewendeten Informationssuchstrategien mehr möglich. Der Prozess der Informationssuche ist für die Analyse des Entscheidungsverhaltens in unseren Experimenten grundsätzlich von grosser Bedeutung. Deshalb ziehen wir es im Allgemeinen vor, unsere Experimente in künstliche Experimental-Szenarien einzubetten, die zwar durchaus einen Bezug zu real existierenden (natürlichen) Umwelten haben können, aber insgesamt so angelegt sind, dass kein bestehendes, in einer natürlichen Umwelt erworbenes Vorwissen zur Lösung der Entscheidungsaufgaben eingesetzt werden kann. Soll trotzdem ein möglichst realitätsbezogenes (d.h. auf realen Daten und Objekten beruhendes) Szenario für das Experiment verwendet werden, so kann man die Anwendung von Vorwissen grösstenteils dadurch ausschliessen, dass man sich auf eine natürliche Umwelt beschränkt, mit der sich die Versuchspersonen nicht auskennen, d.h. keine oder nur sehr beschränkte Erfahrung gemacht haben.

Genau dies haben wir im Quizshow-Szenario getan, das ausführlich in Teil 2 dieser Arbeit beschrieben wird. Das Szenario ist in Anlehnung an die erfolgreiche TV-Quizshow „Wer wird Millionär?“ entstanden. Dadurch, dass wir die Versuchspersonen als Kandidaten einer Quizshow mit vorwiegend äusserst schwierigen Wissensfragen konfrontieren, stellen wir sicher, dass sie zu deren Beantwortung praktisch kein Vorwissen einsetzen können. Das Quizshow-Szenario ermöglicht, den Prozess der Informationssuche als Baustein einer Urteilsheuristik auf individueller Ebene genauer zu erforschen. Die Grundlage hierfür schufen wir durch die (bis dato erstmalige) Verwendung personenbezogener Cues, die uns erlaubt, in jedem Durchgang neue Cues zu präsentieren und damit sowohl die Validitäten als auch die Diskriminationsraten von Durchgang zu Durchgang frei zu variieren⁸. Dank dieser Massnahme war es uns möglich, eine grosse Anzahl

⁸ Die fünf zur Auswahl stehenden Cues werden nach jedem Durchgang durch fünf neue Cues ersetzt ohne dass dabei die Plausibilität des Szenarios beeinträchtigt würde. Als Cues fungieren so genannte Telefon-Joker. Das sind telefonisch kontaktierbare Informationsquellen die zu der gestellten Wissensfrage einen Antwort-Tipp abgeben können. Sie zeichnen sich durch ihre unterschiedliche telefonische Erreichbarkeit (Diskriminationsrate, D) aber auch ihren unterschiedlichen Expertisegrad (Validität, V) aus.

geeigneter Cue-Sets (d.h. Gruppen à fünf Telefon-Joker) zu generieren. Ein Cue-Set ist dann geeignet, wenn sich die hinsichtlich der verschiedenen Gütemasse (V, D, V+D, V*D, SUC) ergebenden Rangfolgen der Cues (absteigend sortiert nach der Cuegüte) möglichst stark – insbesondere am Anfang der Rangfolge – voneinander unterscheiden. Das ist von zentraler Bedeutung für die Analyse des Informationssuchverhaltens, da die sequentielle Wahl der Joker (d.h. die Reihenfolge, in der die Cues konsultiert werden) Aufschluss über die verfolgte Suchstrategie und das zur Beurteilung des Cue-Nutzens herangezogene Güte-Kriterium geben kann. Damit konnten wir das bis anhin bestehende Problem lösen, dass aufgrund der „Suchspur“ d.h. der aufgezeichneten Informationssuch-Reihenfolge in vielen Fällen nicht eindeutig auf ein von der Versuchsperson präferiertes Guegüte-Mass geschlossen werden konnte. Das lag daran, dass sich die hinsichtlich der verschiedenen Gütemasse ergebenden Rangfolgen der Cues in vielen Fällen überlagerten – was insbesondere dann ein Problem war, wenn diese Überlagerung bereits am Anfang der Reihenfolge auftraten (also wenn bspw. als erstes Cue 5 konsultiert wurde und dieser Cue sowohl die höchste Validität als auch den höchsten Success besass).

In Alltagssituationen muss die Güte probabilistischer Informationsquellen jeweils in einer konkreten Umwelt durch praktische Erfahrung gelernt werden. Das Konzept der Cue-Güte beim probabilistischen Entscheiden setzt als psychologisches Modell für menschliches Verhalten voraus, dass der Zusammenhang zwischen einem Cue und einer Zielvariable aufgrund wiederholter Beobachtungen korrekt eingeschätzt werden kann. Um solche erfahrungsabhängige Lernvorgänge genauer untersuchen zu können, haben wir mit dem Äffchen-Szenario eine Experimental-Umgebung entwickelt, in der die Validitäten und Diskriminationsraten der Cues nicht einfach vorgegeben, sondern simultan, während einer Trainingsphase erlernt werden müssen. In späteren Aufgaben müssen diese Cues dann spezifisch aufgrund ihrer Validität, ihrer Diskriminationsrate oder ihrem Success eingesetzt werden. Das Äffchen-Szenario wird in Teil 3 dieser Arbeit ausführlich beschrieben.

Beim Entscheiden spielen Informationen eine wichtige Rolle. Diese sind aber nicht immer top-aktuell, sondern zum Teil bereits älteren Datums, was Entscheidungen auf dieser Basis unsicher macht: Je geringer die zeitliche Stabilität eines Cues und einer Zielvariable, desto unzuverlässiger die Information. Diesen fundamentalen Zusammenhang haben wir in ein Experimentaldesign eingebaut, welches sich – in Anlehnung an bereits bestehende Untersuchungen – an einem Börsenszenario mit Zweialternativenfall orientiert. Hier lassen sich die zeitliche Stabilität von Cues und Zielvariablen sehr gut als zusätzliche Elemente realisieren. Mit der Einführung der Zeitdimension in die Entscheidungsforschung taucht jedoch auch eine neue Ressource auf – Zeit selbst. Diese unterscheidet sich in ihren grundlegenden Eigenschaften von einer monetären Ressource: Sie vergeht und sie kann nicht vermehrt werden. Aus diesem Grund haben wir mit dem Trinkwasser-Szenario ein Experimental-Design entwickelt, in dem Zeit zur zentralen Ressource wird. Dieses Szenario ist bezüglich seiner statischen Umwelt gleich aufgebaut wie das Börsen-Szenario und deshalb hinsichtlich der Resultate mit diesem vergleichbar. Die beiden Szenarien sind im letzten Teil dieser Arbeit (Teil 4) ausführlich beschrieben.

Jeder der vier Teile dieser Arbeit umfasst zwei bis drei Kapitel. Im ersten Kapitel eines jeden Teils wird das in diesem Teil behandelte Experimental-Szenario aus technischer Sicht erläutert. In den ein bis zwei nachfolgenden Kapiteln werden dann konkrete Experimente vorgestellt, die auf dem im ersten Kapitel beschriebenen Szenario beruhen. Anhand dieser Experimente soll exemplarisch illustriert werden, wie mit den verschiedenen Experimental-Szenarien gearbeitet werden kann. Es wird ausführlich dargelegt, welche Fragestellungen untersucht wurden, welche Methodik zum Einsatz kam und welche Resultate wir gefunden haben. Zu sämtlichen in dieser Arbeit vorgestellten Experimenten und Experimental-Szenarien liegen Forschungsberichte der Angewandten Kognitionspsychologie der Universität Zürich (AKZ-Forschungsberichte) vor. Dabei entspricht ein Kapitel dieser Arbeit jeweils dem Inhalt eines AKZ-Forschungsberichts.

Die in dieser Arbeit vorgestellten Experimente sind allesamt Produkte einer teilweise sehr engen Zusammenarbeit verschiedener Personen, die zusammen den Kern der Entscheidungspsychologie am Fachbereich Angewandte Kognitionspsychologie der Universität Zürich (UZH) bildeten. Neben Damian Läge, meinem Doktor-Vater, bei dem ich mich speziell für die gute Betreuung dieser Arbeit bedanke und der immer ein offenes Ohr für Probleme hatte, waren Seraphina Zurbriggen und Daniel Hausmann Teil dieses Kerns. Allen dreien danke ich für die stets angenehme und oft intensive Zusammenarbeit, ohne die diese Arbeit nicht zu Stande gekommen wäre. Entsprechend heterogen ist die Liste Erstautoren bei den oben erwähnten AKZ-Forschungsberichten: Damian (Kapitel 3 und 5), Daniel (Kapitel 6) und Seraphina (Kapitel 10 & 11) danke ich ganz herzlich für ihre Beiträge. Sie haben damit wesentlich zum Gelingen dieser Arbeit beigetragen.

Damian, dessen Übersicht es zu verdanken ist, dass aus den vielen Puzzleteilen letztlich ein vernünftig strukturiertes Ganzes geworden ist, liess mir in den vergangenen Jahren viel Freiheit bei der Wahl meines Arbeitsschwerpunktes. Dadurch hatte ich die Möglichkeit, zusätzlich zu meiner Arbeit an dieser Dissertation, als IT- und Netzwerk-Koordinator für mehrere Fachbereiche am Psychologischen Institut der Universität Zürich (UZH) zu arbeiten und so der Informatik, die mich seit meinen Jugendjahren fasziniert, weiterhin sehr konkret verbunden zu bleiben. Die stark praxisbezogene IT-Koordinationsarbeit bot mir einen idealen Ausgleich zur zuweilen sehr theoretischen Grundlagenforschung in der Entscheidungspsychologie. Nicht zufällig hat sich der inhaltliche Schwerpunkt meiner Arbeit im Verlauf der Zeit von der umfassenden Ausarbeitung und Realisation der Experimente, der Datenerhebung und der inhaltlichen Auswertung der Resultate immer stärker zur Entwicklung der Experimentaldesigns, der programmiertechnischen Realisierung (inklusive Aufbau der Datenbanken), der Präparation der Datenbasis sowie der Ausarbeitung und Umsetzung der benötigten Auswertungsprozeduren hin verlagert. Insbesondere die Zusammenstellung (und Ausbalancierung) der Ausgangsdaten war jeweils mit einem grossen zeitlichen und gedanklichen Aufwand verbunden, da die Datensätze aufgrund der spezifischen Fragestellungen i.d.R. speziellen Ansprüchen hinsichtlich der Auswertbarkeit genügen mussten. Mit anderen Worten: Die Ausgangsdaten mussten jeweils so zusammengestellt werden, dass sie – im Zusammenspiel mit den erhobenen Daten – geeignete Auswertungen (bzw. potentiell aussagekräftige Resultate) bezüglich der untersuchten Fragestellungen ermöglichten.

Die oben erwähnte Verlagerung meines Arbeitsschwerpunkts auf die in erster Linie technischen Aspekte der Experimente war insbesondere deshalb möglich, weil Seraphina, mit der ich erfreulicherweise fast drei Jahre lang das Büro teilen durfte, froh war, sich nicht um die Programmierung unserer gemeinsamen sowie ihrer eigenen Experimente kümmern zu müssen und diesen Teil der Arbeit jeweils gerne mir überliess. Im Gegenzug hat Seraphina die beiden im vierten Teil vorgestellten Experimente geplant, die Datenerhebung geleitet und die Auswertungen durchgeführt. Vergleichbar gestaltete sich die Zusammenarbeit mit Daniel bei dem in Teil 2 (Quizshow-Szenario) präsentierten zweiten Experiment (Kapitel 6). Auch diesem Experiment liegt das von mir programmierte Experimental-Framework zu Grunde. Das Experiment als solches wurde hingegen von Daniel geplant, durchgeführt und ausgewertet. Davon abgesehen, durfte ich im Rahmen unserer Zusammenarbeit sowohl für Daniel als auch für Seraphina noch eine Reihe weiterer Experimente programmieren. Eine besondere Freude und Herausforderung war dabei jeweils der Entwurf möglichst übersichtlich strukturierter und gut handhabbarer Benutzeroberflächen.

Zu guter Letzt möchte ich mich an dieser Stelle beim Schweizerischen Nationalfonds für die Finanzierung der im Rahmen dieser Arbeit durchgeführten Grundlagenforschung bedanken. Ohne die finanzielle Unterstützung durch den SNF wäre diese Arbeit nicht zu Stande gekommen.

Zürich, im August 2008

Stephan Christen

Kapitel 1

Einleitung

1.1 Das Informationboard

Um eine Entscheidung fällen zu können, braucht es Information. Diese Information ist in vielen Fällen nicht vorgegeben, sondern muss zunächst gesucht werden. Der Prozess der Informationsbeschaffung rückt deswegen immer stärker in den Fokus der Entscheidungspsychologie: Nicht nur die Informationslage selbst, sondern auch ihre Genese ist von grosser Bedeutung für die Richtung, in die später eine Entscheidung getroffen wird.

Mit dem Mouselab haben Payne, Bettman & Johnson (1988) ein computergestütztes Tool vorgestellt, mit dem die Phase der Informationsbeschaffung während Entscheidungsprozessen besser beobachtet werden konnte. Dabei sass eine Versuchsperson vor einem Bildschirm, auf dem eine Reihe von Entscheidungsaufgaben präsentiert wurde, welche es zu lösen galt. Zur Interaktion mit dem Experiment diente eine Computermouse – daher der Name Mouselab. Das Mouselab zeichnete im Hintergrund sämtliche Interaktionen der Versuchsperson mit dem Computerprogramm auf. Diese Methode wird auch als „process tracing technique“ bezeichnet. Basierend auf dem Mouselab-Ansatz sind seither viele computergestützte Entscheidungsexperimente entstanden, in deren Zentrum jeweils ein so genanntes Informationboard steht (bspw. Bröder, 2000b; 2005; Newell & Shanks, 2003; Newell, Weston & Shanks, 2003; etc.).

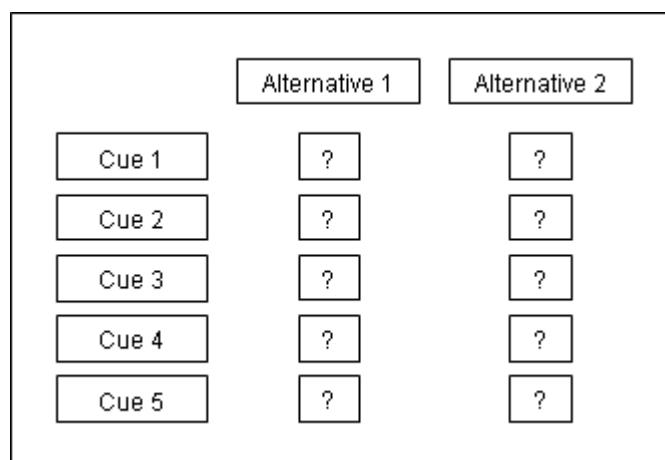


Abbildung 1.1: Beispiel für ein rudimentäres Informationboard mit einem 5x2-Design (5 Cues x 2 Alternativen)

Das Informationboard kann in seiner grundlegendsten Form als eine Art Tafel betrachtet werden, die eine zweidimensionale Matrix enthält (siehe Abbildung 1.1). Entlang der einen Achse (hier der Ordinate) sind verschiedene Informationsquellen (Cues) angeordnet, entlang der anderen Achse verschiedene Entscheidungsalternativen. Die Experimente bestehen aus sehr vielen gleichartigen Durchgängen, in denen es jeweils darum geht, aus den angebotenen Alternativen die bezüglich eines vorgegebenen Kriteriums (tatsächlich) beste auszuwählen, wobei von Runde zu Runde jeweils neue Alternativen angeboten werden. Da eine Versuchsperson über die zur Auswahl stehenden Alternativen i.d.R. kein Vorwissen besitzt, muss sie in jeder Runde wieder

neu (und unter Aufwendung einer Ressource) nach Hinweisen suchen, anhand derer sie die Qualität der Alternativen möglichst gut einschätzen kann. Zu diesen Hinweisen gelangt sie, indem sie im Informationboard mit dem Mauszeiger auf die als Schaltflächen ausgestalteten Cues klickt, worauf (ebenfalls im Informationboard) für jede Alternative ein Hinweis angezeigt wird⁹. Im Hintergrund protokolliert das Computerprogramm mit, welche Cues in welcher Reihenfolge angeklickt werden, welche Hinweise die Versuchsperson dabei erhält und welche Entscheidung sie schliesslich trifft. Von Interesse können ggf. auch die folgenden process tracing Variablen sein: Die Anzahl der konsultierten Cues, die Anzahl der gefundenen Informationen, oder die für die Aufgaben benötigten Bearbeitungszeiten. Abhängig vom Gegenstand der Untersuchung sind viele dieser Aufzeichnungen von essentieller Bedeutung für die spätere Analyse und Interpretation des Such- und Entscheidungsverhaltens der Versuchsperson.

1.2 Die drei Building-Blocks: Such-, Stopp- und Entscheidungsregel

In Entscheidungsprozessen ist irgendwann der Moment erreicht, in dem eine Person die Informationssuche beendet und eine Entscheidung trifft. Der Abbruch der Informationssuche erfolgt zu meist, weil die Person der Meinung ist, genügend Informationen gefunden zu haben, um eine gute Entscheidung treffen zu können, oder (seltener) weil ihr keine Cues oder keine weiteren Ressourcen (wie Geld oder Zeit) mehr für die weitere Suche zur Verfügung stehen. Bei der Analyse solch relativ einfacher und schnell ablaufender Entscheidungsprozesse wird üblicherweise zwischen drei separaten Phasen, den so genannten building blocks unterschieden (Gigerenzer, Todd, & the ABC Research Group, 1999). Diese drei building blocks sind die Such-, die Stopp- und die Entscheidungsregel. Sie beschreiben, wie nach Informationen gesucht wird, wann die Suche abgebrochen wird und wie anhand der gefundenen Informationen schliesslich eine Entscheidung getroffen wird. Der Prototyp solcher einfacher Entscheidungs-Verfahren ist die Take-the-Best-Heuristik (TTB) von Gigerenzer, Hoffrage & Kleinbölting (1991). Gemäss dieser Strategie werden die zur Verfügung stehenden Cues in der Reihenfolge ihrer Validität konsultiert und die Informationssuche beendet, sobald eine der Informationsquellen eine Entscheidungsgrundlage offenbart. Die Entscheidung wird anschliessend lediglich auf der Basis dieser einen Information getroffen. Eine Entscheidung nach TTB wird deshalb auch als ein „One-Reason Decision Making“ (ORDM) bezeichnet (Gigerenzer & Goldstein, 1999, S. 79). Dem ORDM stehen die Entscheidungsstrategien des More-Reason Decision Making (MRDM) gegenüber, bei dem Entscheidungen auf der Grundlage der Integration mehrerer Informationen gefällt werden. Die meisten im Folgenden zu beschreibenden Experimental-Szenarien lassen in der Regel sowohl ORDM als auch MRDM zu.

⁹ Man spricht in diesem Zusammenhang auch von einer attributweisen Informationssuche, weil man sich die Cues als Attribute der zur Auswahl stehenden Alternativen vorstellen kann. In Experimenten, in denen die Hinweise nicht nur attribut- sondern auch alternativenweise abgerufen werden können, sind auch die Alternativen als Schaltflächen ausgestaltet. In diesem Fall erhält man von allen vorhandenen Cues gleichzeitig je einen Hinweis zur angeklickten Alternative. Möchte man, dass die Hinweise eines Cues für jede einzelne Alternative separat abgerufen werden können, kann man bei n Cues und m Alternativen $n \cdot m$ separate Schaltflächen im Informationboard anbieten.

1.3 Ursprüngliche Experimental-Designs zur Untersuchung von TTB

Mit ihrem vielzitierten Städte-Experiment konnte die Gruppe um Gigerenzer den Nachweis erbringen, dass man mit der TTB-Heuristik ebenso gute Entscheidungen treffen kann wie mit einer vollständigen Analyse aller vorliegenden Cue-Informationen (Czerlinski, Gigerenzer & Goldstein, 1999; Gigerenzer & Goldstein, 1996; 1999; Martignon & Hoffrage, 1999). Da es sich bei diesem Nachweis um eine Simulationsstudie handelte, war damit noch nicht gezeigt, dass die TTB-Heuristik von Personen in der Realität auch tatsächlich angewendet wird. Die Studie hat jedoch eine experimentelle Forschung inspiriert, die sich der Klärung dieser Frage widmete. So knüpfen die ersten Informationboards zur Informationssuche beim Entscheiden entweder an das Städteparadigma von Gigerenzer an oder aber an Finanzmarkt-Entscheidungen. Letzteres wird vor dem Hintergrund der traditionell engen Verknüpfung der Entscheidungsforschung mit ökonomischen Theorien verständlich.

1.3.1 Informationssuche in der Umwelt

TTB wurde ursprünglich als Heuristik für Informationssuche im Gedächtnis entworfen (Gigerenzer et al., 1991). Es dauerte allerdings einige Jahre, bis die Anwendung von TTB als Gedächtnisabrufstrategie erstmals experimentell untersucht wurde (Bröder & Schiffer, 2003). Demgegenüber steht die Informationssuche in der Umwelt. Dort lassen sich die drei building blocks von TTB genau so gut anwenden und können experimentell wesentlich leichter beobachtet werden.

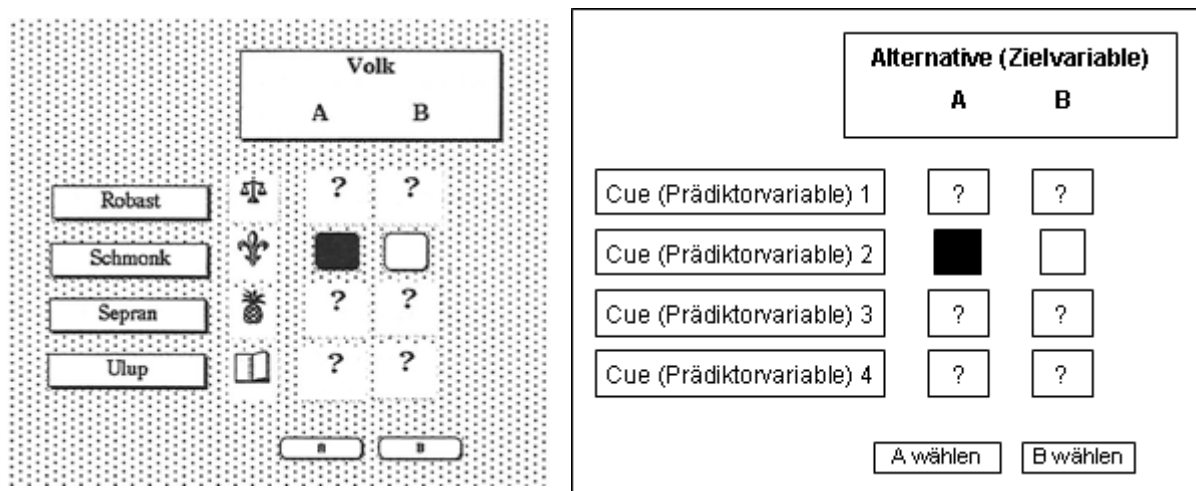


Abbildung 1.2: Das von Bröder (2000b, Experiment 1) verwendete Informationboard mit einem 4x2-Design. Original (links, entnommen aus Bröder, 2000b) und schematische Darstellung (rechts)

Da TTB spezifische Vorhersagen über die einzelnen Phasen des Entscheidungsprozesses (die building blocks) macht, lassen sich diese empirisch gut überprüfen. Entsprechende Experimente wurden erstmals von Bröder durchgeführt (vgl. Bröder, 2000b, Experimente 1-3 & 5). Seine Experimente zum Cue-Lernen (Validitätslernen) basieren auf einem Informationboard mit einem 4x2- bzw. 5x2-Design – d.h. mit 4 bzw. 5 Cues und jeweils 2 Alternativen (vgl. Abbildung 1.2).

In einem künstlichen Szenario mussten die Versuchspersonen den Zusammenhang zwischen der Grösse einiger (fiktiver) Völker und verschiedener (fiktiver) Kulturtechniken, die von den Völkern entweder beherrscht werden oder nicht, gelernt werden. Das Informationboard diente hier lediglich der Präsentation der Cue-Ausprägungen in der Lernphase: Die Ausprägungen der Cues wurden der Reihe nach für jeden der vier Cues einzeln präsentiert, bis insgesamt 50 Lerndurchgänge pro Cue erreicht waren. Die anhand der bisherigen Lerndurchgänge feststellbaren Cue-Validitäten wurden zur „Optimierung des Häufigkeitslernens“ (Bröder, 2000b, S.94) nach jedem Durchgang in der Form von Balkendiagrammen rückgemeldet. In einer „diagnostischen Phase“ wurde anschliessend der Lernerfolg überprüft: Dazu mussten die Versuchspersonen die geschätzten Cue-Validitäten auf einer Skala von 0.5 bis 1.0 angeben und anschliessend mittels paarweiser Vergleiche angeben, welchen der beiden präsentierten Cues sie für den valideren halten. (Sowohl Bröder als auch andere Autoren – vgl. Bröder & Schiffer, 2003; Newell & Shanks, 2003; Newell, Weston & Shanks, 2003; Newell, Rakow, Weston & Shanks, 2003; sowie Hausmann, 2004 – konnten mit ihren Studien aufzeigen, dass nicht alle Versuchspersonen TTB anwenden.)

Das Informationboard zeigt die Cues für zwei Aktien (Aktie 1 und Aktie 2) in einer 4x2-Matrix. Die Cues sind: Aktientrend positiv?, Kapitalreserven?, Investitionen?, etablierte Firma?. Die Ausprägungen sind: nein, ja, ?, ?. Die Aktionäre können entweder 1 Information kaufen oder alle Informationen kaufen. Die Wahl zwischen Aktie 1 und Aktie 2 ist unten zu sehen. Rechts ist der aktuelle Stand des Spielers angegeben: Bereits investiert -200 \$, Penunzen.

	Aktie 1	Aktie 2
Aktientrend positiv?	nein	ja
Kapitalreserven?	nein	ja
Investitionen?	?	?
etablierte Firma?	?	?

Bereits investiert -200 \$
Penunzen

1 Information kaufen
Alle Informationen kaufen

Wähle:
Aktie 1 Aktie 2

Abbildung 1.3: Informationboard im Börsenspiel von Bröder (2000b, Experiment 6). Entnommen aus Bröder (2000b).

Auch für sein Experiment zur Untersuchung von Kontexteffekten (wie z.B. Suchkosten) auf den Prozess der Informationssuche verwendete Bröder ein 4x2-Design (Bröder, 2000b, Experiment 6). Im Gegensatz zu den vorangegangenen Experimenten beinhaltete Experiment 6 kein Validitätslernen. Die Informationen mussten aktiv in der Umwelt gesucht werden, wobei die Cues nur in einer fix vorgegebenen Reihenfolge, nämlich absteigend nach ihren Validitäten geordnet, abgerufen werden konnten. Die Experimental-Umgebung hat Bröder als Börsenspiel ausgestaltet, bei dem pro Durchgang zwei verschiedene Aktien zur Auswahl standen (siehe Abbildung 1.3).

Beide Aktien waren durch binäre Werte auf den vier angebotenen Cues charakterisiert. Die Aufgabe bestand darin, jeweils die Aktie mit der höheren Gewinnchance auszuwählen.

Um den Einfluss von Persönlichkeitsvariablen (d.h. von individuellen Präferenzen) auf die angewendete Suchregel zu messen, setzte Bröder für seine nächsten Studien wiederum das Börsenspiel-Szenario ein (siehe Bröder, 2005, S. 109ff; vgl. Bröder 2000b, Experiment 6), das später auch von Ben Newell et al. übernommen wurde (Newell & Shanks, 2003; Newell, Weston & Shanks, 2003). Das Ziel von Experiment 1 (Bröder, 2005) bestand darin, die in früheren Studien aufgetretenen individuellen Unterschiede in der Wahl der Suchstrategie auf konsistente Persönlichkeitseigenschaften zurückzuführen. Zu diesem Zweck war die Informationsstruktur so angelegt, dass für die Anwendung von TTB und Franklin's Rule (bei der die verschiedenen Cues bzw. Attribute differentiell gewichtet werden) ein etwa gleich hoher erwarteter Gewinn resultiert, so dass allfällige Unterschiede in der verwendeten Strategiewahl vornehmlich auf persönliche Präferenzen und nicht auf die experimentelle Umwelt zurückgeführt werden können. Im Unterschied zum Experiment, das der Untersuchung von Kontexteffekten diente, konnten die Cue-Informationen in Experiment 1 in beliebiger Reihenfolge abgerufen werden.

Dieselbe Oberfläche verwendete Bröder auch in Experiment 2 (Bröder, 2005), mit dem er den Einfluss von Adaptivität, Intelligenz und Lernen auf die Strategiewahl untersuchte. In einer der Versuchsbedingungen war der eigentlichen Anwendungsphase eine aus 45 Durchgängen bestehende Lernphase vorgeschaltet. In dieser wurden sämtliche Cue-Informationen offen im Informationboard präsentiert. Für die Experimente 3 bis 6 benutzte Bröder (2005) im Wesentlichen identische Designs, wobei in diesen jeweils 3 Entscheidungsalternativen (Aktien) zur Auswahl standen, statt wie bis anhin nur deren 2. Zudem wurden in den Experimenten 4a bis 4c Massnahmen zur Erhöhung der Belastung des Arbeitsgedächtnisses ergriffen. So erhielten die Versuchspersonen bspw. in Experiment 4a anstelle binär ausgeprägter Cue-Hinweise jeweils eine Prozentangabe. Dieses komplexere Informationsformat musste mittels eines kognitiven Rekodierungsprozesses zuerst in eine nicht-numerische Aussage (Aktien-Bewertung) transformiert werden, bevor dann auf der Grundlage dieser (transformierten) Information eine Entscheidung gefällt werden konnte.

Um herauszufinden, welche Rolle (mal-)adaptive Routinetendenzen in Entscheidungssituationen spielen, kleidete Bröder die Informationsstruktur aus der ersten Bedingung von Experiment 6 für die zweite Bedingung in eine neue Umwelt (ein Immobilienfonds-Szenario) ein. Die beiden Umwelten (Börsenszenario in Bedingung 1 resp. Immobilienfonds in Bedingung 2) waren absolut strukturgleich aufgebaut. Sie unterschieden sich lediglich in ihrer semantischen Einkleidung: Das Informationboard wurde für Bedingung 2 um 90 Grad gedreht, so dass die Attribute dort zeilenweise angeordnet waren, und auch der Bildschirmhintergrund war anders gestaltet. Ausserdem standen in Bedingung 2 anstelle von Aktien nun verschiedene Fonds zur Auswahl, und auch die Cues hatten andere Bezeichnungen.

Newell und seine Kollegen haben für ihre Studien zur Untersuchung der Bedingungen, unter denen Personen TTB anwenden, das 4x2-Börsendesign von Bröder (2000b) übernommen (Ne-

well, Weston & Shanks, 2003). Auch bei ihnen mussten die Versuchspersonen in einer Reihe von „forced-choice“-Entscheidungen jeweils die bessere von zwei angebotenen Aktien auswählen (siehe Abbildung 1.4).

	Share A	Share B
Share trend positive?	NO	YES
Financial reserves?	YES	YES
Invest in new projects?	NO	YES
Established company?	YES	YES
	CHOOSE SHARE A	CHOOSE SHARE B
Probability that Chosen Share will be most Profitable	<input type="text"/>	
Most Profitable Share	<input type="text"/>	
Private Account	<input type="text"/>	

NEXT TRIAL

Abbildung 1.4: Informationboard im Börsenspiel von Newell & Shanks (2003). Entnommen aus Newell & Shanks (2003).

Den insgesamt drei Experimenten war jeweils eine Lernphase vorgeschaltet, in der sämtliche Cue-Informationen offen vorlagen und die Validitätsfolge der Cues gelernt werden musste. Mit Experiment 1 wurde der Einfluss unterschiedlich hoher Suchkosten untersucht. Die Kosten pro Information betrugen in beiden Bedingungen konstant 1 Penny. Abhängig von der Bedingung konnten die Versuchspersonen entweder 5 Pence (Bedingung mit hohen Informationskosten) oder 10 Pence (Bedingung mit tiefen Informationskosten) gewinnen. In der ersten Bedingung betrugen die Kosten pro Information somit 1/5 des möglichen Gewinns, in der zweiten Bedingung lediglich 1/10. Experiment 2 war eine Replikation von Experiment 1 – mit drei Ausnahmen: In der Lernphase erhielten die Versuchspersonen sowohl nach der Hälfte der Durchgänge als auch vor Beginn der Anwendungsphase Hinweise („hints“), mit deren Hilfe die korrekte Validitätsfolge der Cues erschlossen werden konnte. Ausserdem fiel Bedingung 2 mit den tiefen Informationskosten weg, und am Ende des Experiments mussten die Versuchspersonen in einem Fragebogen Auskunft über die in der Lernphase verwendeten Strategien geben. Experiment 3 war ähnlich aufgebaut wie Experiment 2. Allerdings wurde die stochastische Umgebung durch eine deterministische ersetzt. Dort war immer diejenige Alternative tatsächlich die bessere, für die (aufgrund der Cue-Informationen) mittels Bayes-Theorem berechnet eine höhere a-posteriori Wahrscheinlichkeit resultierte. In der zweiten Bedingung von Experiment 3 wurden die Cues

und deren Ausprägungen semantisch neu eingekleidet („Distinct Labels“), um eine simple Auszählstrategie (aufsummieren von „Yes“- bzw. „No“-Antworten) zu erschweren.

Das Börsendesign von Bröder wurde auch für die Studie zu Success (Newell, Rakow, Weston & Shanks, 2003) und die Studie zur Stopppregel (Newell, Weston & Shanks, 2003) verwendet. In der Studie zur Untersuchung der Stopppregel kamen allerdings eine 6x2-Variante (Experiment 1) und eine 2x2-Variante (Experiment 2) zum Einsatz. Experiment 1 umfasste 180 Aufgaben. Die Versuchspersonen erhielten nach 60 und nach 120 Aufgaben einen „hint“, der die korrekte Reihenfolge der Nützlichkeit (Validität) der beiden Cues verriet. Eine Information kostete 1 Penny und für eine korrekte Entscheidung gab es 7 Pence zu gewinnen. In Experiment 2, das insgesamt 120 Durchgänge umfasste, erhielten die Versuchspersonen bereits nach 30 Durchgängen einen Hinweis auf die relative Nützlichkeit (Validität) der Cues. Eine Information kostete wiederum 1 Penny und eine korrekte Entscheidung brachte 5 Pence ein.

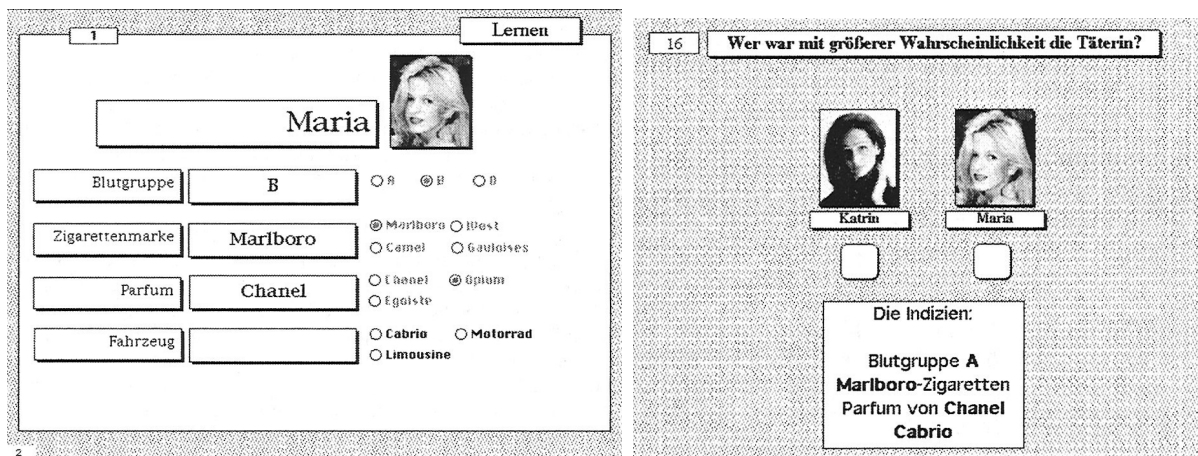


Abbildung 1.5: Informationboard während der Lernphase (links) und Bildschirmoberfläche während der Anwendungsphase (rechts) im Börsenspiel von Bröder (2005, Experiment 7). Entnommen aus Bröder (2005).

1.3.2 Informationssuche im Gedächtnis

Spezifischere Studien zum Cue-Lernen (siehe Bröder, 2005, Experimente 7 bis 11) erforderten ein neuartiges Versuchsdesign. In den Experimenten 7 bis 10 untersucht Bröder erstmalig probabilistische Inferenzen aus dem Gedächtnis, mit dem Ziel, unterschiedliche Informationsintegrationsstrategien zu identifizieren. Ausgangspunkt von Bröders Studien ist die von Gigerenzer & Todd (1999) vorgebrachte Gedächtnisabruf-Hypothese, die postuliert, dass Entscheidungsprozesse, die mit vorgegebener Information arbeiten, sich fundamental von Entscheidungsprozessen unterscheiden sollten, die mit gedächtnisbasierter Information arbeiten, weil die Informationssuche im Gedächtnis Ressourcen beansprucht. Zur Untersuchung der Gedächtnisabruf-Hypothese verwendete Bröder eine ähnliche experimentelle Logik wie Hoffrage et al. (2000) in einer Studie zum Rückschaufehler: Bei Hoffrage et al. mussten die Versuchspersonen in einer Lernphase je drei Cue-Werte für eine Reihe von Objekten auswendig lernen und damit gewissermaßen ein mentales Abbild einer fiktiven Umwelt aufbauen. Später mussten sie dann für Paare dieser Ob-

jekte Schlussfolgerungen bezüglich einer vorgegebenen Zielvariablen ziehen. Bei Bröder mussten die Versuchspersonen in der Lernphase eines fiktiven Experimental-Szenarios insgesamt 10 Personenprofile lernen (siehe Abbildung 1.5, links). Jedes Profil umfasste dieselben 4 Attribute (Blutgruppe, Zigarettenmarke, Parfum, Autotyp), von denen jedes 3 verschiedene Ausprägungen annehmen konnte. In der darauf folgenden Phase wurden in jedem Durchgang die Namen und Portrait-Photos von jeweils zwei dieser Personen – die eines Mordes verdächtigt wurden – präsentiert. Anhand von vier Indizien¹⁰ mussten die Versuchspersonen dann entscheiden, welche der beiden Verdächtigen mit grösserer Wahrscheinlichkeit die Täterin sei (siehe Abbildung 1.5, rechts). Hierzu mussten die Indizien (unter Berücksichtigung ihrer Relevanz) gedächtnisbasiert mit den gelernten Personenprofilen verglichen werden. Die Validitätshierarchie der Attribute (Cues) wurde mittels verschiedener Hinweise eingeführt. So wurde bspw. angegeben, dass das gefundene Blut mit sehr hoher Wahrscheinlichkeit von der Täterin stamme, etc. Das Experiment umfasste zwei Bedingungen: Eine mit hoher und eine mit tiefer Belastung des Arbeitsgedächtnisses. Erwartet wurde, dass ein substantieller Teil der Vpn eine TTB-Heuristik anwenden würde, sofern die Hypothese gilt, dass Gedächtnisabruf kognitive Kosten verursacht. Die Resultate schienen die Hypothese zu bestätigen. Aufgrund eines methodischen Problems führte Bröder später Experiment 8 durch, in dem er Informationsstruktur aus Experiment 7 übernahm, jedoch die Cues neu einkleidete, um einen Materialeffekt ausschliessen zu können (siehe Abbildung 1.6, links). Ausserdem variierte er die Belastung des Arbeitsgedächtnisses über die beiden Bedingungen: Im Gegensatz zur Bedingung mit hoher Belastung wurde in der Bedingung mit tiefer Belastung während der Anwendungsphase nicht nur die Namen und Portrait-Fotos der beiden Verdächtigen präsentiert, sondern deren vollständige Profile (mit den konkreten Ausprägungen auf allen vier Attributen; siehe Abbildung 1.6, rechts). Die gefundenen Resultate stützen die Gedächtnisabruf-Hypothese.

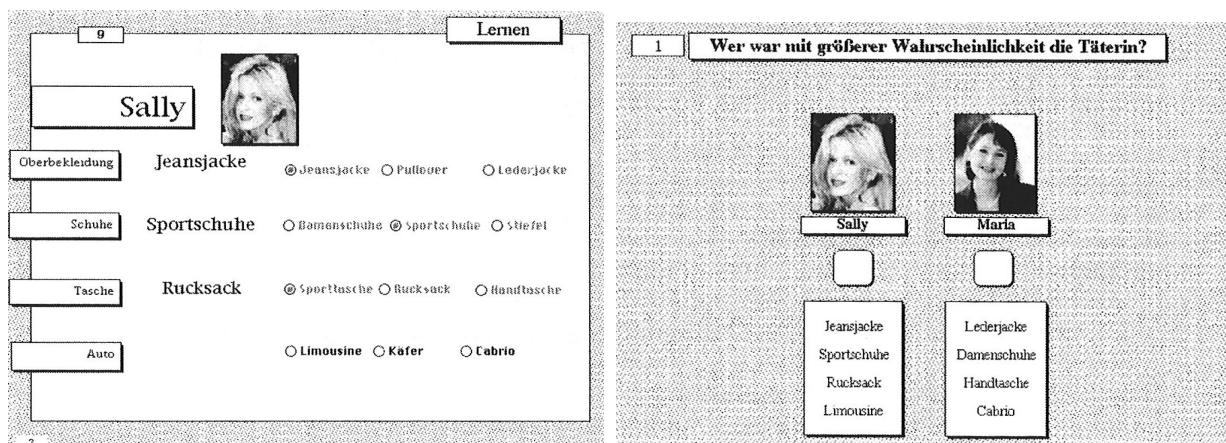


Abbildung 1.6: Informationboard während der Lernphase (links) und Bildschirmoberfläche während der Anwendungsphase (rechts) im Börsenspiel von Bröder (2005, Experiment 8). Entnommen aus Bröder (2005).

¹⁰ Für jedes Attribut wurde eine konkrete Ausprägung angegeben, bspw.: „Blutgruppe A“, „Marlboro-Zigaretten“, „Parfum von Chanel“, „Cabrio“.

Experiment 9 diente der Überprüfung der Format-Hypothese, die besagt, dass ein bildliches internes Repräsentationsformat die Abrufkosten senkt, da ein holistischer Informationsabruf erfolgt. Aus diesem Grund wurde in einer der beiden Bedingungen versucht, eine bildliche interne Enkodierung des Lernmaterials zu erreichen (siehe Abbildung 1.7, links). Die Lernphase der verbalen Bedingung entsprach exakt derjenigen in Experiment 8. Ein Format-Effekt konnte in diesem Experiment nicht gefunden werden. Um die bildliche Enkodierung der Stimuli zu fördern, realisierte Bröder in Experiment 10 neben Bedingung 1, in der die Ausprägungen der Attribute verbal enkodiert waren, zwei weitere Bedingungen. In der „Bild-Bedingung“ wurde das Lernmaterial bildlich dargeboten und zusätzlich bei jedem Personenprofil ein ganzheitliches Bild der Person angezeigt (siehe Abbildung 1.7, rechts). In der Entscheidungsphase der Bild+Prototyp-Bedingung wurde zusätzlich ein anhand der Indizien erstelltes Prototypenbild der Täterin präsentiert. Diese Massnahmen sollten den Aufwand für den Informationsabruf reduzieren (d.h. eine Verringerung der Suchkosten bewirken) und die Häufigkeit von TTB über die drei Bedingungen damit abnehmen. Im Gegensatz zu Experiment 9 scheinen die Resultate von Experiment 10 die Format-Hypothese zu bestätigen.

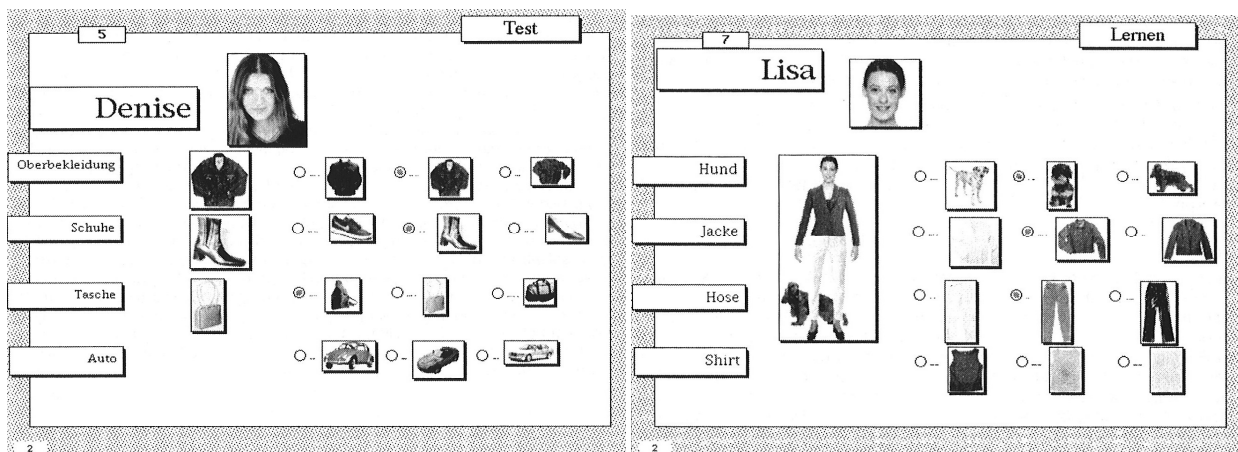


Abbildung 1.7: Informationboards während der Lernphasen der Experimente 9 (links) und 10 (rechts) bei Bröder (2005, Experiment 9 & 10). Entnommen aus Bröder (2005).

Aufgrund der Resultate aus den Experimenten 7-10 zieht Bröder das vorläufige Fazit, dass die Cue-basierte Informationssuche tatsächlich mit relativ hohen subjektiven Kosten verbunden zu sein scheint und dass deshalb auf ein sparsames One-Reason Decision-Making zurückgegriffen wird (vgl. Bröder 2005, S.221).

Die Frage nach der Einkleidung einer Fragestellung in ein Experimental-Setting stellte sich auch für Lee & Cummins (2004). Sie kritisierten an den bisherigen Studien, dass die darin benutzten Referenzklassen („stimulus domains“) jeweils künstlich konstruiert worden sind, obwohl Gigerenzer & Todd (1999) auf die Bedeutung einer strukturierten Umwelt für das effiziente Funktionieren von schnellen und sparsamen Heuristiken („fast-and-frugal heuristics“) hingewiesen haben. Aus diesem Grund benutzten Lee & Cummins für ihre Studie eine Referenzklasse, die auf

„natürlichen“ Originaldaten (Daten über die Galapagos-Inseln) beruhte, wobei sie aus methodischen Gründen einige Stimuli (Alternativen) entfernten und einige Cues komplementär verwenden mussten. Die Tatsache, dass die Daten aus einer natürlichen Umwelt stammen, musste vor den Versuchspersonen verschleiert werden, um zu verhindern, dass interindividuelle Unterschiede im Wissen über die Galapagosinseln das Verhalten der Versuchspersonen beeinflussen. Die Stimuli wurden deshalb nicht in ihrer natürlichen Form repräsentiert, sondern hinter einer künstlichen Oberfläche versteckt. Die durch die Stimuli definierte Informations-Struktur (Cue-Struktur) war von dieser Änderung nicht betroffen und entsprach derjenigen der natürlichen Umwelt. In der neuen Repräsentationsform wurde jeder Stimulus als (transparenter) Gaskanister auf dem Bildschirm dargestellt. In jedem Kanister waren verschiedenfarbige, sich bewegende Gasmoleküle zu sehen. Als Zielvariable musste die relative Giftigkeit der Gase beurteilt werden, wobei die verschiedenen Farben der Gasmoleküle mit den einzelnen Cues korrespondierten. Die Gasmoleküle bewegten sich innerhalb des Kanisters zufällig umher, um zu verhindern, dass konfigurale Effekte durch die Überlagerung verschiedener Moleküle entstehen. Es implizierte ausserdem, dass zwischen den Cues keine festen räumlichen Beziehungen bestehen. Nach einer Trainingsphase, in der die Cue-Validitäten erworben wurden, folgte zur Überprüfung des Cue-Lernens eine Phase mit paarweisen Vergleichen zwischen den verschiedenen Gasen. Nach jeder Entscheidung musste die Antwortsicherheit auf einer 5-stufigen Skala angegeben werden. Feedback über die Korrektheit der Entscheidung wurde nicht gegeben.

In Experimenten, in denen kein Validitäts-Lernen stattfindet, können die Validitäten der Cues ggf. direkt im Informationboard angegeben werden. Da die Suche nach Information in der Umwelt immer unter Aufwendung einer Ressource erfolgt, berücksichtigen Newell, Rakow, Weston & Shanks (2003) sowie Läge, Hausmann, Christen & Daub (2005) in ihren Experimenten neben der Validität auch die Diskriminationsrate der Cues. In drei der vier nachfolgend präsentierten Experimental-Umgebungen wird die Diskriminationsrate konsequent berücksichtigt. In einer der drei Experimental-Umgebungen müssen die Validitäten und Diskriminationsraten der Cues erfahrungsbasiert erlernt und anschliessend gedächtnisbasiert angewendet werden. In den anderen beiden Experimental-Umgebungen werden sowohl die Validitäten der Cues als auch deren Diskriminationsraten in prozentualer Form im Informationboard angegeben.

1.4 Weiterentwicklung des Informationboard anhand spezifischer Fragestellungen

1.4.1 Die Fragestellungen

Die nachfolgend aufgeführten Fragestellungen weiten die Forschung zur Informationssuche beim Entscheiden aus und bedingen die Entwicklung neuer Experimental-Umgebungen. Diese basieren aus experimentalmethodischer Sicht alle auf dem klassischen Informationboard-Ansatz, der jeweils spezifisch an die zu klärenden Fragestellungen adaptiert wurde.

1. *Aktive Informationssuche* – Welche Bedeutung hat die aktive Informationssuche für die spätere Entscheidung? Entscheiden sich Personen anders, wenn sie aktiv nach Informationen suchen

müssen, als wenn der Prozess der Informationssuche entfällt, weil bereits alle (relevanten) Informationen offen vorliegen? (Optionsschein-Szenario).

2. *Informationssuchstrategien (Suchregeln)* – Welche Strategien verfolgen Personen bei der sequentiellen Suche nach Informationen? Welche Kriterien ziehen sie zur Beurteilung der Cuegüte heran? (Quizshow-Szenario).

3. *Cue-Lernen* – Sind Personen in der Lage, Cues spezifisch aufgrund ihrer Validität (V), ihrer Diskriminationsrate (D) oder ihrem Success einzusetzen, wenn V und D nicht vorgegeben werden, sondern in einer Lernphase simultan gelernt werden müssen? Was genau wird dabei gelernt? (Äffchen-Szenario).

4. *Zeitliche Stabilität von Cues und Zielvariablen / Zeit als Ressource*: Welchen Einfluss haben Zeit als Ressource sowie zeitlich instabile Cues und Zielvariablen auf den Prozess der Informationssuche? Wie gehen Personen mit zeitlich instabilen Informationen um und wie reagieren sie auf Zeit als Ressource? (Trinkwasser-&-Börsen-Szenario).

In den folgenden Unterkapiteln wird dargelegt, warum wir zur Untersuchung der oben aufgeführten Fragestellungen neue Experimental-Umgebungen entwickeln mussten. Es wird geschildert, wie die Fragestellungen in neue Szenarien umgesetzt wurden und welche Innovationen damit verbunden sind.

1.4.2 Das Optionsschein-Szenario - Die Bedeutung aktiver Informationssuche

Im Alltag ist die Suche nach Information eine zentrale Voraussetzung für viele Entscheidungen. In theoretischen Modellen wird die Informationssuche deshalb häufig als Phase des Entscheidungsprozesses betrachtet. In klassischen Entscheidungsexperimenten wurde die aktive Informationssuche jedoch meistens ausgeklammert und stattdessen sämtliche relevante Information offen vorgegeben (siehe bspw. Bröder, 2000b). Vor diesem Hintergrund haben wir das Optionsschein-Szenario entwickelt, mit dem wir die Bedeutung der aktiven Informationssuche für die spätere Entscheidung untersuchen können (siehe Christen, Hausmann & Läge, 2007; vgl. Läge, Christen & Hausmann, 2007). Das Optionsschein-Szenario ist die Implementierung eines einfachen Börsenspiels, das in Anlehnung an das Börsenspiel von Bröder (2000b) entwickelt wurde. Im Unterschied zum Börsenspiel von Bröder können die Versuchspersonen in diesem Szenario einen Geldbetrag auf bis zu fünf verschiedene Optionsscheine setzen. Als Informationsquellen stehen fünf Prädiktorvariablen (so genannte „Indikatoren“) zur Verfügung, die für jeden einzelnen Optionsschein einen Hinweis auf einen Wertzuwachs oder Wertverlust geben können. Das Szenario bietet zwei Versuchsbedingungen mit identischer Informationsstruktur im Hintergrund. Anhand dieser kann beobachtet werden, ob sich Personen systematisch anders entscheiden, wenn die Informationssuche entfällt, weil bereits alle Informationen offen vorliegen (Bedingung 1, siehe Abbildung 2.1), als wenn sie aktiv nach Informationen suchen müssen (Bedingung 2, siehe Abbildung 2.2, die den Initialzustand für diese Bedingung zeigt, sowie Abbildung 2.9, die das Informationboard zeigt, nachdem eine Versuchsperson bereits zwei Mal nach Informationen gesucht hat).

In der ersten Bedingung zeigen die insgesamt fünf probabilistischen Cues in jeder Spielrunde sämtliche Informationen (Ausprägungen) für die fünf Optionsscheine offen an, ohne dass die Versuchsperson zuvor danach hat suchen oder eine Ressource hat aufwenden müssen. In der zweiten Bedingung müssen diese Informationen hingegen aktiv gesucht und sequentiell aufgedeckt werden¹¹, wobei das Aufdecken nur unter Aufwendung einer Ressource möglich ist (d.h. es wird jeweils ein bestimmter monetärer Betrag vom Spielgeldkonto abgezogen). Mittels Vergleich des Entscheidungsverhaltens zwischen den beiden Bedingungen messen wir die Entscheidungsvarianz, die aufgrund der identischen Informationsstrukturen der beiden Bedingungen nur auf den Faktor Informationssuche zurückgeführt werden kann.

1.4.3 Das Quizshow-Szenario – Unterscheidung von Informationssuch-Strategien mit flexibel anbietbaren Validitäten und Diskriminationsraten

Im Zentrum des Quizshow-Szenarios steht die Frage nach den Suchregeln, die Individuen bei der Informationssuche anwenden. Mit Hilfe dieses Szenarios soll geklärt werden, anhand welcher Kriterien (Validität, Diskriminationsrate, $V+D$, $V*D$, Success) Personen die Güte von Cues beurteilen. Es geht also darum, die Suchregel als Baustein einer Urteilsheuristik auf individueller Ebene genauer zu erforschen.

Ein methodisches Problem früherer Experimente bestand in der Schwierigkeit der empirischen Differenzierung zwischen TTB und linearen Modellen aufgrund überlappender Vorhersagen (siehe bspw. Bröder 2000a, 2000b; Hoffrage et al., 2000; Martignon & Hoffrage, 2002). Bröder (2005, S.63) spricht in diesem Zusammenhang vom „Separierungsproblem“. Um besser zwischen den angewendeten Strategien differenzieren zu können, erhöhte er in seinen Experimenten (Bröder, 2000b) die Anzahl der angebotenen (dichtotomen) Cues von 3 auf 4, was in neun von 120 möglichen Paarvergleichen der Cue-Muster zu unterschiedlichen Vorhersagen von TTB und Franklin’s Rule führte. Wenn wir innerhalb von TTB zwischen der Anwendung unterschiedlicher Suchregeln differenzieren wollen, stehen wir vor einem vergleichbaren Separierungsproblem¹²: In den bisher durchgeführten Experimenten standen jeweils nur relativ wenige verschiedene Cues zur Auswahl, so dass nur eine sehr begrenzte Anzahl unterschiedlicher Cue-Sets (= Cue-Gruppen) gebildet werden konnte und von diesen wenigen Cue-Sets viele keine gute Differenzierung zwischen den theoretisch postulierten Suchregeln (V , D , $V+D$, $V*D$, Success)

¹¹ Aufgrund der Tatsache, dass die Angabe der konkreten Ausprägung einer Prädiktorvariablen von der Variablen-Deklaration getrennt werden kann, lässt sich für Bedingung 2 auf einfache Weise eine an die gedächtnisbasierte Cue-Anwendung angelehnte Informationssuch-Phase modellieren. Dabei zeigt das den Versuchspersonen präsentierte Informationboard zwar von Anfang an die fünf Prädiktorvariablen an, nicht aber deren konkrete Ausprägungen. Diese müssen aktiv gesucht bzw. zur Anzeige gebracht werden. Dies geschieht, indem die dargebotenen Prädiktorvariablen resp. die zur Auswahl stehenden Optionsscheine mit dem Mauszeiger einzeln angeklickt werden. Die Informationssuche kann nach Belieben entweder nur alternativenweise oder nur attributweise oder sowohl alternativenweise als auch attributweise erfolgen.

¹² Anhand der „Suchspur“ d.h. der aufgezeichneten Informationssuch-Reihenfolge konnten wir bisher nicht eindeutig auf ein von der Versuchsperson präferiertes Güte-Mass schliessen, weil sich die hinsichtlich der verschiedenen Gütemasse ergebenden normativen Rangfolgen der Cues in vielen Fällen überlagerten. Das war insbesondere dann ein Problem, wenn diese Überlagerung bereits am Anfang der Reihenfolge auftrat.

erlaubten. In einem künstlichen Experimental-Szenario ohne Bezug zu einer realen (natürlichen) Umwelt hätte man das Problem einfach durch die Einführung sehr vieler fiktiver Cues lösen können, da man deren Validitäten und Diskriminationsraten aufgrund des fehlenden Realitätsbezugs beliebig hätte festlegen können. In einer künstlichen Versuchsumgebung riskiert man allerdings, dass die Versuchspersonen schnell ihre Motivation verlieren und das Experiment zu einer reinen Klick-Übung verkommt. Mit einem realitätsnahen Szenario lässt sich dieses Problem zwar etwas entschärfen, steht dafür aber vor dem Problem, dass sich zu einer vorgegebenen Zielvariablen meistens nur eine Hand voll inhaltlich plausibler und voneinander unabhängiger Cues finden lässt. Ein weiteres Problem das bei einem natürlichen Szenario auftaucht ist das von Bröder (2005, S. 63) geschilderte „Identifizierungsproblem“. Es bezeichnet die Tatsache, dass Rückschlüsse auf die genaue Form der Informationsintegration nur möglich sind, wenn die Wissensbasis der Person bekannt ist. Da aufgrund der geschilderten Probleme einem künstlichen Szenario möglichst aus dem Weg zu gehen ist, wurde mit der Quizshow eine andere Lösung entwickelt, um die Validitäten und Diskriminationsraten der Cues von Durchgang zu Durchgang variieren zu können.

Die Lösung für das eben geschilderte Problem liegt in der bis dato erstmaligen Verwendung personenbasierter Cues, die als Telefon-Joker in ein fiktives Quizshow-Szenario eingebettet sind (siehe Christen, Läge & Hausmann, 2007; vgl. Hausmann, Christen & Läge, 2005, sowie Läge, Hausmann, Christen & Daub, 2005). In diesem Szenario wird den Versuchspersonen eine Reihe sehr schwieriger Wissensfragen gestellt, bei deren Beantwortung die telefonisch kontaktierbaren Experten (Telefon-Joker) behilflich sein können. Die Cue-Validitäten lassen sich als unterschiedliche Expertisegrade¹³ der Telefon-Joker realisieren, und die Diskriminationsraten als deren unterschiedliche telefonische Erreichbarkeiten (siehe Abbildung 4.28). Auf diese Weise können V und D innerhalb der zulässigen Bereiche beliebig und unabhängig voneinander variiert werden, ohne die Plausibilität des Szenarios damit zu gefährden. Obwohl bei der Erzeugung der Cues und der Zusammenstellung der Cue-Sets einige einschränkende Kriterien zu berücksichtigen sind, ist es dennoch möglich, ausreichend viele unterschiedliche Cue-Sets zusammenzustellen, die sich zur Differenzierung der fokussierten Suchregeln (V, D, V+D, V*D, SUC) eignen¹⁴.

Im Quizshow-Szenario sind die Prädiktorvariablen also keine klassischen probabilistischen Cues mehr, die man als die Attribute einer Zielvariablen bezeichnen könnte und die in einer direkten „ökologischen“ Relation zu dieser Zielvariablen stehen. Stattdessen handelt es sich bei diesen nun um menschliche Informationsquellen, die Tipps abgeben. Bei diesen Tipps können wir inhaltlich nicht mehr von „Ausprägungen“ sprechen. Für unsere Zwecke können probabilistische Cues und personifizierte Informationsquellen jedoch als formal identisch angesehen werden: In

¹³ Der Expertisegrad eines Telefon-Jokers entspricht der Wahrscheinlichkeit, mit der er die gestellte Wissensfrage korrekt beantworten wird. Wir sprechen deshalb mitunter auch von der „Zuverlässigkeit“ eines Jokers.

¹⁴ Ein Cue-Set ist dann geeignet, wenn sich die hinsichtlich der verschiedenen Gütemasse (V, D, V+D, V*D, SUC) ergebenden Rangfolgen der Cues (absteigend sortiert nach der Cuegüte) möglichst stark – insbesondere am Anfang der Rangfolge – voneinander unterscheiden. Das ist von zentraler Bedeutung für die Analyse des Informationssuchverhaltens, da die Reihenfolge in der die Cues konsultiert werden, Aufschluss über die verfolgte Suchstrategie und das zur Beurteilung des Cue-Nutzens herangezogene Güte-Kriterium geben kann.

beiden Fällen „konsultiert“ man „Informationsquellen“, die probabilistische Hinweise liefern, anhand derer man dann auf die Ausprägungen der Zielvariablen schliessen kann.

Eine weitere im Quizshow-Szenario implementierte Neuerung ist das Zwei-Phasen-Design zur Analyse der verwendeten Suchregeln, wobei Phase 1 der Hypothesen-Generierung und Phase 2 der Hypothesen-Überprüfung dient. In Phase 1 differenzieren viele der Cue-Sets noch nicht eindeutig zwischen allen fünf Suchregeln. Die Gesamtmenge der Cue-Sets in Phase 1 ist jedoch so zusammengestellt, dass wir damit insgesamt gut zwischen den fünf fokussierten Suchregeln differenzieren können. In Phase 2 wird sodann überprüft, ob eine der beiden in Phase 1 am häufigsten detektierten Suchregeln klar als die von der Versuchsperson präferierte Suchstrategie identifiziert werden kann. In der zweiten Phase sind die Cue-Sets denn auch so präpariert, dass mit jedem einzelnen Cue-Set exklusiv entweder die eine oder die andere der beiden überprüften Suchregeln detektiert wird. So findet also in Phase 1 eine Eingrenzung von fünf auf zwei der theoretisch postulierten Suchstrategien statt, und in Phase 2 werden diese beiden Strategien dann exklusiv gegeneinander getestet (siehe Abbildung 4.1). Der Knackpunkt bei diesem Zwei-Phasen-Design liegt in der Zusammenstellung einer ausreichenden Anzahl geeigneter Cue-Sets, unter der Einschränkung der Anzahl Aufgaben die noch in vernünftiger Zeit gelöst werden können. So stehen in Phase 1 lediglich 50 Aufgaben bzw. Cue-Sets für die Hypothesen-Generierung zur Verfügung, die zudem hinsichtlich der Detektierbarkeit der fünf fokussierten Suchregeln fair ausbalanciert sind.

1.4.4 Das Äffchen-Szenario – Erlernbarkeit von Cue-Qualitäten

In der PMM-Theorie beschreiben Gigerenzer et. al. (1991), wie die interne Repräsentation von Cue-Validitäten zu Stande kommt: „[...] cue validities are learned by observing the frequencies of co-occurrences in an environment“ (S. 510). Objekte werden also im Alltag zu Cues, weil wir andauernd Zusammenhänge zwischen dem Vorhandensein (resp. der Abwesenheit) dieser Objekte und anderen, gleichzeitig auftretenden (resp. fehlenden) Objekten lernen. Wir bilden, überprüfen und verwerfen täglich, ob bewusst oder unbewusst, eine grosse Anzahl Zusammenhangs-Hypothesen. Sofern nicht bewusst ein Zusammenhang zwischen den wahrgenommenen Objekten hergestellt wird, ist möglicherweise eine relativ enge raum-zeitliche Koinzidenz im Auftreten bzw. in der Wahrnehmung dieser Objekte nötig, damit ein allfälliger Zusammenhang unbewusst gelernt werden kann. Durch die wiederholte Beobachtung des gleichzeitigen Auftretens zweier Objekte lernen wir nicht nur, dass zwischen dem Auftreten dieser Objekte ein (allenfalls gerichteter) Zusammenhang besteht – wodurch ein Objekt zum Hinweis auf ein anderes Objekt wird (und umgekehrt) – sondern wir lernen ein Gefühl für die Häufigkeit bzw. die Qualität dieses (gerichteten) Zusammenhangs gleich mit. Dieser Prozess, bei dem die Güte probabilistischer Informationsquellen durch praktische Erfahrung gelernt werden muss, wird als Cue-Lernen bezeichnet.

Einigen der ersten Experimente zur Überprüfung von TTB mit dem einfachen Mouselab-Setting war eine Lernphase vorgeschaltet, in der die Versuchspersonen zunächst durch Testaufgaben die Validität der einzelnen Cues herausfinden sollten (z.B. Bröder, 2000b, Experimente 1-5; Newell & Shanks, 2003; Lee & Cummins, 2004; Hausmann, Christen & Läge, 2005). Auf diese Art

konnte nur sporadisch hinreichendes Lernen der Cue-Validität festgestellt werden. Mangels ausreichender Lernerfolge wurden die tatsächlichen Validitäten den Versuchspersonen dann zum Teil angegeben. So blenden bspw. Newell & Shanks (2003, Experimente 2&3) während der Lernphase gelegentlich eine Liste mit den korrekten Validitäten ein und Bröder (2000b) gibt in seinem Experiment in jedem Durchgang für jeden Cue in Balkenform die relativen Häufigkeiten der bis dato korrekten und falschen Vorhersagen an. Dadurch kann nicht mehr eindeutig festgestellt werden, was die Versuchspersonen tatsächlich durch Beobachtung gelernt haben und was sie sich aus den Vorgaben gemerkt haben. Andere Experimente, wie bspw. dasjenige von Lee & Cumins (2004), scheitern allein schon aufgrund einer ungeeigneten Experimentalanordnung an der Erbringung eines klaren Nachweises dafür, dass Validitätsfolgen gut gelernt werden können. Im Experiment von Newell, Rakow, Weston & Shanks (2003) musste zusätzlich zur Validität auch Diskriminationsrate und Success gelernt werden. Hier konnte lediglich aufgrund des Verhaltens in der anschliessenden Runde ermittelt werden, ob die Personen eher gemäss Validität oder gemäss Success Informationen suchten. Es zeigte sich, dass „weder ... noch“ die beste Beschreibung des Verhaltens war, woraus unter Annahme eines rationalen Entscheidungsverhaltens zu schliessen wäre, dass die meisten Versuchspersonen keinen realistischen Eindruck der Cue-Qualitäten bekommen haben. Man muss all diesen Experimenten zugute halten, dass Cue-Lernen hier nur als Vorbedingung für das eigentlich die Forscher interessierende Entscheidungsverhalten galt und nie direkt im Fokus stand.

Vor diesem Hintergrund bildet das Äffchen-Szenario (siehe Christen, Hausmann & Läge, 2008; vgl. Christen, Hausmann & Läge, 2006) einen Weg, sich in der experimentellen Forschung auf das Erlernen der Cue-Qualitäten zu konzentrieren.

In dieser Experimental-Umgebung müssen die Validitäten und Diskriminationsraten der Cues simultan während einer Trainingsphase erlernt werden. In der zweiten Phase des Experiments (der „Anwendungsphase“) müssen diese Cues dann aufgabenspezifisch aufgrund ihrer Validität, ihrer Diskriminationsrate oder ihrem Success eingesetzt werden. Neu an dieser Experimental-Umgebung ist einerseits die Oberfläche: Das Experiment ist in ein medizinisches Szenario eingekleidet, in dem Äffchen gerettet werden müssen, die von einer tödlichen Krankheit bedroht sind. Ausserdem werden die Cues nicht mehr in Listenform präsentiert, sondern sind auf einer gedachten Ellipse angeordnet, um möglichst keine bestimmte Cuerangfolge vorzugeben (siehe Abbildung 7.7a). Andererseits, und das ist das Entscheidende, ist die Informationsstruktur im Hintergrund so gestaltet, dass wir damit erstmals eindeutig feststellen können, ob die Versuchspersonen während der Lernphase die Validitäten und/oder Diskriminationsraten lernen oder gleich das integrierte Gütemass Success. Die Dauer der Lernphase ist adaptiv: Je schneller eine Versuchspersonen ein adäquates Gefühl für die Güte der Cues entwickelt, desto kürzer dauert die Lernphase. Mindestens hundert Lerndurchgänge sind aber in jedem Fall zu absolvieren. Hinweise auf die tatsächlichen Validitäten und Diskriminationsraten der Cues werden keine gegeben – weder während der Lernphase noch danach. Der Fokus liegt bei diesem Szenario ganz klar auf der Untersuchung des Cue-Lernens, ohne dass dabei Kompromisse hinsichtlich anderer interessierender Fragestellungen an die Experimentalanordnung gemacht werden müssten. Entscheidend ist, dass die während des Cue-Lernens aufgebaute subjektive Repräsentation der Cue-Güte

die Basis für das Entscheidungsverhalten einer Person darstellt. Deshalb muss TTB hinsichtlich dieser subjektiven Repräsentationen auf individuellem Versuchspersonenniveau getestet werden können. Im Unterschied zu den Studien von Bröder (2000a; 2000b), Newell & Shanks (2003) und Newell, Weston & Shanks (2003), die das Entscheidungsverhalten auf Stichprobenebene untersuchten, erlaubt das Äffchen-Szenario die Auswertungen individueller Verhaltensmuster hinsichtlich der Such-, Stopp-, und Entscheidungsregeln. Zudem können damit individuelle Lernunterschiede berücksichtigt werden.

1.4.5 Das Trinkwasser-&-Börsen-Szenario – Zeitliche Stabilität von Cues & Zielvariablen / Zeit als Ressource

Bis anhin wurde die Informationssuche nur bei unabhängigen Einzelereignissen untersucht, ohne dabei die Zeitdimension zu berücksichtigen. Die Informationssuche fand dabei jeweils innerhalb einer stabilen Umwelt statt. Um probabilistische mentale Modelle (nach Gigerenzer et al., 1999) an sich verändernde Umwelten und dynamische Cue-Landschaften anbinden zu können, wurde eine neue Experimental-Umgebung – das Trinkwasser-&-Börsen-Szenario – entwickelt (Christen, Zurbriggen, Hausmann & Läge, 2007).

Im Zentrum dieses Doppel-Szenarios steht die Frage, ob Personen sensitiv auf unterschiedliche (stabile vs. dynamische) statistische Umwelten reagieren. Welche Rolle spielt es, ob Serien von Entscheidungen mit oder ohne zeitlichen Kontext getroffen werden? Können Personen bei Entscheidungen, die in einem Zeitfluss miteinander verbunden sind, die zeitliche Dimension und die Informationen über die zeitliche Stabilität eines Systems in der Informationsnutzung berücksichtigen? Allgemeiner formuliert: Sind Personen überhaupt in der Lage, bei ihren Entscheidungen Informationen aus der Vergangenheit zu berücksichtigen und dabei gleichzeitig auf deren „Verfallsquote“ zu achten? Ausserdem lässt sich untersuchen, ob Personen Informationen nur für die Gegenwart suchen oder auch den prospektiven Nutzen einer Information in ihr Suchverhalten einbeziehen.

Auch diese Experimental-Umgebung verwendet ein Informationboard zur Messung des Informationssuch- und Entscheidungsverhaltens, um auf die individuell angewendeten Such-, Stopp- und Entscheidungsregeln schliessen zu können. Die entscheidende Neuerung gegenüber bisherigen Experimental-Umgebungen im Bereich der Informationssuche ist die, dass die Versuchspersonen das gesamte Experiment nicht als eine Reihe von Einzeldurchgängen erleben, sondern in einen Zeitfluss eintauchen.

Als Informationen über die zeitliche Stabilität des Systems wurden zwei neue strukturelle Parameter eingeführt: Erstens S_Z (die Stabilitätsrate einer Entscheidungs-Alternative Z, kurz „Zielvariablen-Stabilität“) und zweitens S_I (die Stabilitätsrate einer Informationsquelle I, kurz „Cue-Stabilität“). Mit S_I liegt damit ein (neben V und D) drittes Mass zur Beurteilung der Cue-Güte vor. Ein (hoch valider) Cue mit einer geringen zeitlichen Stabilität liefert Informationen, auf die man sich bereits nach kurzer Zeit nicht mehr verlassen kann. Ein zeitlich stabiler (und ebenfalls hoch valider) Cue liefert hingegen Informationen, die auch nach mehreren (im Experimentaldesign ca. zwei bis drei) Zeitschritten noch immer eine relativ hohe Validität aufweisen. Ähnliches

gilt für die Stabilität einer Zielvariablen (S_z): Die Ausprägung einer zeitlich stabilen Zielvariablen bleibt über einen längeren Zeitraum hinweg konstant, während dies für eine zeitlich weniger stabile Zielvariable nicht gilt.¹⁵

Mit dem Börsen-Szenario (siehe Abbildung 9.9) lässt sich der Einfluss messen, den einerseits das Wissen um die Stabilitäten der Zielvariablen und andererseits das Wissen um die Stabilitäten der Cues auf das Such- und Entscheidungsverhalten im Zwei-Alternativenfall hat. Das Börsen-Szenario unterscheidet sich vor allem aufgrund der dahinter liegenden Daten, aber auch auf der Benutzeroberfläche relativ stark vom Optionsschein-Szenario (s.o.), da es auf völlig andere Fragestellungen fokussiert: Zum einen wird im Börsen-Szenario die Diskriminationsrate vernachlässigt. Zudem sind dort die Informationen nur einzeln abrufbar und es stehen nur noch 2 Alternativen zur Auswahl.

So wie das Börsen-Szenario wurde auch das Trinkwasser-Szenario (siehe Abbildung 9.3) komplett neu entwickelt. Seine Oberfläche unterscheidet sich jedoch nur geringfügig von derjenigen des Börsen-Szenarios. Auch im Trinkwasser-Szenario können die Informationen nur einzeln abgerufen werden. Zudem ist seine statistische Umwelt prinzipiell identisch aufgebaut wie die des Börsen-Szenarios. Dadurch können die beiden Szenarien bezüglich der verwendeten unterschiedlichen Ressourcen problemlos miteinander verglichen werden. Mit dem Trinkwasser-Szenario kann man untersuchen, welchen Einfluss die Verwendung unterschiedlicher Ressourcen hat, genauer: Was passiert, wenn die üblicherweise verwendete monetäre Ressource durch die Ressource Zeit ersetzt ist? Da Zeit ein wichtiger Faktor in Entscheidungssituationen ist (vgl. Maule & Edland, 1997, S. 189), soll das Paar an Settings auch zeigen, ob in Abhängigkeit der verwendeten Ressource Unterschiede im Such- und/oder Entscheidungsverhalten auftreten.

1.5 Literatur

- Bröder, A. (2000a). Assessing the empirical validity of the "Take The Best"-heuristic as a model of human probabilistic inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 1332-1346.
- Bröder, A. (2000b). "Take The Best – Ignore The Rest". *Wann entscheiden Menschen begrenzt rational?* Lengerich: Pabst.
- Bröder, A. (2005). *Entscheiden mit der „Adaptiven Werkzeugkiste“: Ein empirisches Forschungsprogramm*. Lengerich: Pabst Science Publishers.
- Bröder, A., & Schiffer, S. (2003). "Take The Best" versus simultaneous feature matching: Probabilistic inferences from memory and effects of representation format. *Journal of Experimental Psychology: General*, 132, 277–293.

¹⁵ Die Stabilitätsrate ist dabei definiert als die Wahrscheinlichkeit, mit der die Ausprägung einer Variablen im nächsten Durchgang (bzw. zum nächsten Messzeitpunkt) noch dieselbe ist wie im aktuellen Durchgang (bzw. zum aktuellen Messzeitpunkt). Die Stabilitätsrate einer Zielvariablen oder eines Cues mit einer zeitlich konstanten Ausprägung beträgt somit 100%. Eine Variable, deren Ausprägung sich von Messzeitpunkt zu Messzeitpunkt ändert, hat hingegen eine Stabilitätsrate von 0%. Die Stabilitätsrate ist folglich kein absoluter Wert, sondern abhängig von der Anzahl der festgelegten Messzeitpunkte und wird dementsprechend als prozentualer Anteilswert (d.h. unter der Massgabe von Beobachtungszeiträumen) angegeben.

- Christen, S., Hausmann, D., & Läge, D. (2006). *Ist ein Gefühl für Informationsgüte lernbar? Ein Lernexperiment zu Validität, Diskriminationsrate und Success von Informationsquellen*. AKZ-Forschungsbericht 37, Zürich: Angewandte Kognitionspsychologie.
- Christen, S., Hausmann, D., & Läge, D. (2007). *Ein Optionsscheinszenario zur Untersuchung der Strategien sequentieller Informationssuche: Aufbau, Ablauf und technische Spezifikation der Experimentalumgebung*. AKZ-Forschungsbericht 42, Zürich: Angewandte Kognitionspsychologie.
- Christen, S., Hausmann, D., & Läge, D. (2008). *Das „Äffchen-Szenario“: Technische Spezifikation einer Experimentalumgebung zur Untersuchung der Lernbarkeit von Validität, Diskriminationsrate und Success von Informationsquellen*. AKZ-Forschungsbericht 61, Zürich: Angewandte Kognitionspsychologie.
- Christen, S., Läge, D., & Hausmann, D. (2007). *Die Quizshow: Ein experimentelles Design zur Messung von individuellen Informationssuch-Strategien in einfachen Urteilsheuristiken*. AKZ-Forschungsbericht 55, Zürich: Angewandte Kognitionspsychologie.
- Christen, S., Zurbriggen, S., Hausmann, D., & Läge, D. (2007). *Trinkwasser- und Börsenszenario: Aufbau, Ablauf und technische Spezifikation einer Experimentalumgebung zur entscheidungsvorbereitenden Informationssuche in zeitlichen Verläufen*. AKZ-Forschungsbericht 43, Zürich: Angewandte Kognitionspsychologie.
- Czerlinski, J., Gigerenzer, G., & Goldstein, D. G. (1999). How good are simple heuristics? In G. Gigerenzer, P. M. Todd & the ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 97–118). New York: Oxford University Press.
- Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, 103, 650–669.
- Gigerenzer, G., & Goldstein, D. G. (1999). Betting on one good reason: The Take The Best heuristic. In G. Gigerenzer, P. M. Todd & the ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 75–95). New York: Oxford University Press.
- Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic Mental Models: A Brunswikan theory of confidence. *Psychological Review*, 98, 506–528.
- Gigerenzer, G., Todd, P. M., et al., Eds. (1999). *Simple Heuristics That Make Us Smart*. New York: Oxford University Press.
- Hausmann, D. (2004). *Informationssuche im Entscheidungsprozess. Die Nützlichkeit von Hinweis-Cues und der Anspruch an Urteilssicherheit*. Dissertationsschrift. Zürich: Zentralstelle der Studentendruckerei.
- Hausmann, D., Christen, S., & Läge, D. (2005). *Und sie lernen es doch! Das Einschätzen der Validität von probabilistischen Hinweis-Cues und der Einfluss der subjektiven Repräsentation auf das Urteilsverhalten*. AKZ-Forschungsbericht 08. Zürich: Angewandte Kognitionspsychologie.
- Hoffrage, U., Hertwig, R., & Gigerenzer, G. (2000). *Hindsight Bias: A by-product of knowledge updating?* *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 26, 566–581.
- Lee, M. D., & Cummins, T. D. R. (2004). Evidence accumulation in decision making: Unifying the “take the best” and the “rational” models. *Psychonomic Bulletin and Review*, 11, 343–352.
- Läge, D., Christen, S., & Hausmann, D. (2007). *Der Einfluss der Informationssuche auf die Entscheidung*. AKZ-Forschungsbericht 04. Zürich: Angewandte Kognitionspsychologie.
- Läge, D., Hausmann, D., Christen, S., & Daub, S. (2005). *Was macht einen "guten Cue" aus? Strategien der Informationssuche beim heuristischen Entscheiden unter Unsicherheit*. AKZ-Forschungsbericht 05. Zürich: Angewandte Kognitionspsychologie.
- Martignon, L., & Hoffrage, U. (1999). Why does one-reason decision making work? A case study in ecological rationality. In G. Gigerenzer, P. M. Todd & the ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 119–140). New York: Oxford University Press.

- Martignon, L., & Hoffrage, U. (2002). Fast, frugal, and fit: Simple heuristics for paired comparison. *Theory & Decision*, 52, 29-71.
- Maule, A. J., & Edland, A. C. (1997). The effects of time pressure on human judgement and decision making. *Decision making: cognitive models and explanations*. R. Ranyard, W. R. Crozier and O. Svenson. London: Routledge.
- Newell, B. R., & Shanks, D. R. (2003). Take the best or look at the rest? Factors influencing "one-reason" decision-making. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 29, 53-65.
- Newell, B. R., Rakow, T., Weston, N. J., & Shanks, D. R. (2003). Search strategies in decision-making: the success of "success". *Journal of Behavioral Decision Making*, 17, 117-137.
- Newell, B. R., Weston, N. J., & Shanks, D. R. (2003). Empirical tests of a fast and frugal heuristic: not everyone "takes-the-best". *Organizational Behavior and Human Decision Processes*, 91, 82-96.
- Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 14, 534-552.
- Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. Cambridge: Cambridge University Press.
- Zurbriggen, S., Christen, S., Hausmann, D., & Läge, D. (2008a). *Informationssuche beim Entscheiden in zeitlich dynamischen Umwelten*. AKZ-Forschungsbericht 44. Zürich: Angewandte Kognitionspsychologie.
- Zurbriggen, S., Christen, S., Hausmann, D., & Läge, D. (2008b). *Adaptivität der Informationssuche in statischen und in dynamischen Umwelten*. AKZ-Forschungsbericht 46. Zürich: Angewandte Kognitionspsychologie.
- Zurbriggen, S., Hausmann, D., Christen, S., & Läge, D. (2007a). *Die zeitliche Dynamik von Umwelten: Heuristisches Wahlverhalten am Beispiel eines Trinkwasser-Szenarios mit und ohne Zeitachse*. AKZ-Forschungsbericht 41. Zürich: Angewandte Kognitionspsychologie.
- Zurbriggen, S., Hausmann, D., Christen, S., & Läge, D. (2007b). *Ressource Zeit statt Geld: Ein Vergleich zweier Szenarien zur Entscheidungsforschung*. AKZ-Forschungsbericht 45. Zürich: Angewandte Kognitionspsychologie.

Teil I - Optionsschein-Szenario

Kapitel 2

Ein Optionsscheinszenario zur Untersuchung der Strategien sequentieller Informationssuche: Aufbau, Ablauf und technische Spezifikation der Experimentalumgebung

Wie wichtig die psychologische Modellierung des Prozesses von Informationssuche in der Entscheidungsforschung ist, konnten Läge, Christen & Hausmann (2007) in einem direkten Vergleich zweier Versuchsbedingungen (alle Information vorgegeben versus Information zunächst aktiv zu suchen) zeigen. Das Experiment basierte auf einem Börsenszenario, in welchem zu fünf Optionsscheinen jeweils fünf probabilistische Cues vorlagen. Dieses Experimentalsdesign wird hier in seinem Aufbau und in seinen technischen Spezifikationen ausführlich beschrieben.

2.1 Einleitung

Für Entscheidungssituationen könnte sich Informationssuche als ein kritischer Faktor erweisen. Möglicherweise müssen Entscheidungssituationen, bei denen sämtliche für die Entscheidung relevanten Informationen bereits vorliegen, und Entscheidungssituationen, bei denen die für die Entscheidung relevanten Informationen zuerst aktiv gesucht werden, als zwei unterschiedliche Situationsklassen betrachtet werden. Das Konzept der Informationssuche wurde in der Entscheidungspsychologie bereits mehrfach diskutiert (z.B. Huber, Wider & Huber, 1997, sowie Huber & Macho, 2001). Uns sind jedoch keine Experimente bekannt, in denen diese beiden Situationsklassen im Hinblick auf das resultierende Entscheidungsverhalten miteinander verglichen werden. Mit dem hier beschriebenen Experiment wollen wir nun genau dies tun.

Das Experiment ist unterteilt in zwei grundsätzlich identische Experimentalbedingungen, die sich lediglich bezüglich der Bereitstellung entscheidungsrelevanter Informationen unterscheiden, wobei die Informationen in beiden Bedingungen auf einem Information Board (vgl. Payne 1976) dargeboten werden: Während in der ersten Experimentalbedingung sämtliche entscheidungsrelevanten Informationen in jedem Versuchsdurchgang von Anfang an offen vorliegen, müssen die Informationen in der zweiten Bedingung aktiv gesucht, d.h. einzeln auf dem Information Board aufgedeckt werden. Die aktive Informationssuche muss dabei ressourcenabhängig sein, da ohne Ressourcen-Einsatz prinzipiell immer alle Informationen aufgedeckt würden bzw. werden könnten und sich die beiden Experimentalbedingungen damit nicht substantiell voneinander unterscheiden würden. In der zweiten Bedingung können Informationen deshalb nur im Austausch gegen einen bestimmten Anteil der zur Verfügung stehenden monetären Ressource (Spielgeld) aufgedeckt werden.

Mittels der beiden Experimentalbedingungen wollen wir herausfinden, ob sich Personen auf der Basis vorgegebener Informationen systematisch anders entscheiden als wenn sie aktiv, d.h. unter Aufwendung einer Ressource, nach Informationen suchen müssen.

Neben dieser Hauptfragestellung wollen wir mit dem Experiment auch überprüfen können, welche der folgenden drei Vorgehensweisen Personen bei der Informationssuche (inklusive Stoppregel) bevorzugen: Das One-Reason Decision Making (ORDM), eine Alternativenprüfung im Sinne des Satisficing oder die Überprüfung des vollständigen Informationsraums (bspw. unter Anwendung einer vereinfachten Integrationsregel wie der Dawes' Rule).

Im Hinblick auf die Entscheidungsregel soll uns das Experiment zudem erlauben herauszufinden, ob Personen eher das beste Argument (wie bei der Take-The-Best Heuristik) oder die Integration der vorliegenden Information (z.B. durch einfaches Abzählen der Argumente wie bei der Tallying-Strategie) präferieren.

Untersuchungen von McDougall (1995) zeigen, dass Risiko-meidende Personen sich über einzelne Alternativen genauer informieren (also zu Informationsintegration tendieren), während Risiko-geneigte Personen möglichst viele Alternativen anhand einiger weniger Merkmale bewerten (also eher zu ORDM tendieren). Wir wollen mit unserem Experiment deshalb auch der Frage

nachgehen, ob ein Zusammenhang zwischen riskantem Entscheidungsverhalten (d.h. alles Geld auf eine einzige Alternative setzen) und einer höheren Tendenz zu Informationssuchstrategien im Sinne von ORDM besteht.

Das hier vorgestellte Experiment (in Form eines Börsenspiels) zur Untersuchung der oben genannten Fragestellungen, wurde in Microsoft Visual Basic 6.0 programmiert und vollständig auf einem PC mit dem Betriebssystem Windows XP durchgeführt. In den nachfolgenden Kapiteln dieses Forschungsberichts werden Aufbau, Ablauf und technische Spezifikationen des Experiments beschrieben. Eine exakte Herleitung der Fragestellungen sowie eine Beschreibung der Ergebnisse findet sich in Läge, Christen & Hausmann (2007).

2.2 Allgemeine Struktur des Experiments

Das Experiment ist als ein Börsenspiel konzipiert und besteht im Kern aus den beiden in ihrer Datenstruktur identischen Experimentalbedingungen B1 und B2: In B1 liegen sämtliche entscheidungsrelevanten Informationen bereits offen auf dem Information Board vor (Abbildung 2.1). In B2 hingegen müssen diese Informationen von den Versuchspersonen aktiv, d.h. unter Ressourcenaufwand, gesucht werden (Abbildung 2.2).

Die beiden Abbildungen geben jeweils den Zustand zu Beginn des ersten Versuchsdurchgangs wieder. Mit B2 können wir somit nicht nur das Entscheidungsverhalten der Versuchspersonen erfassen, so wie in B1, sondern zusätzlich auch deren Informationssuchverhalten.

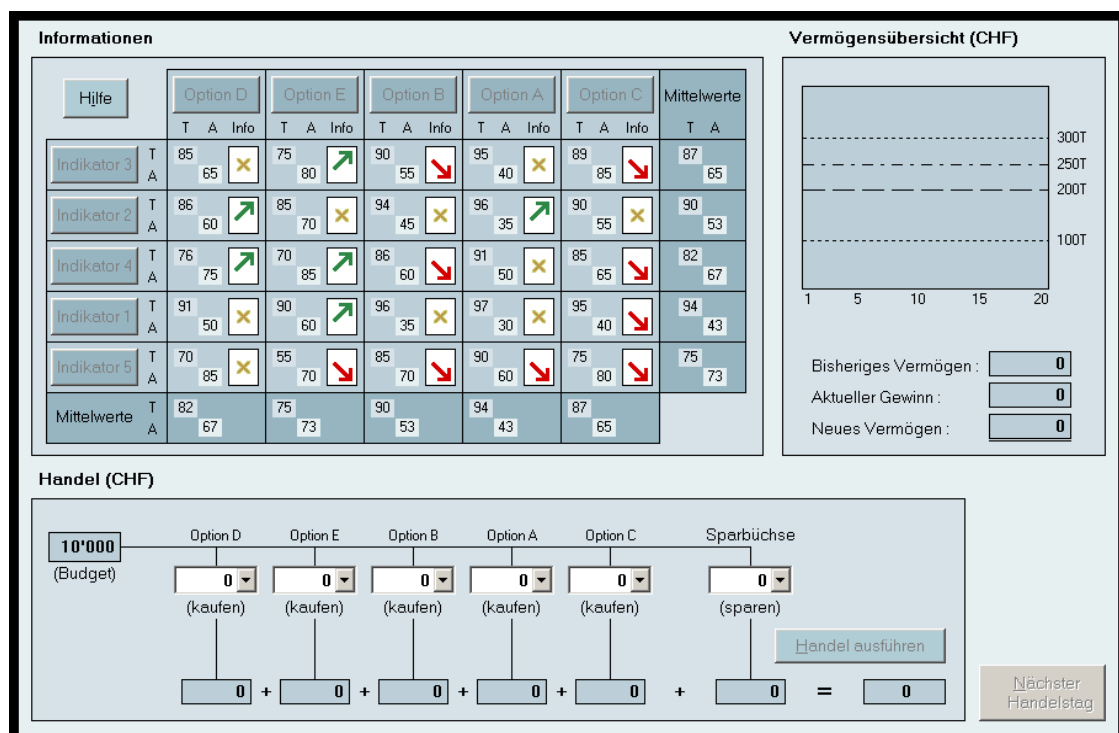


Abbildung 2.1: Bildschirmoberfläche von Experimentalbedingung B1 (Initialzustand)

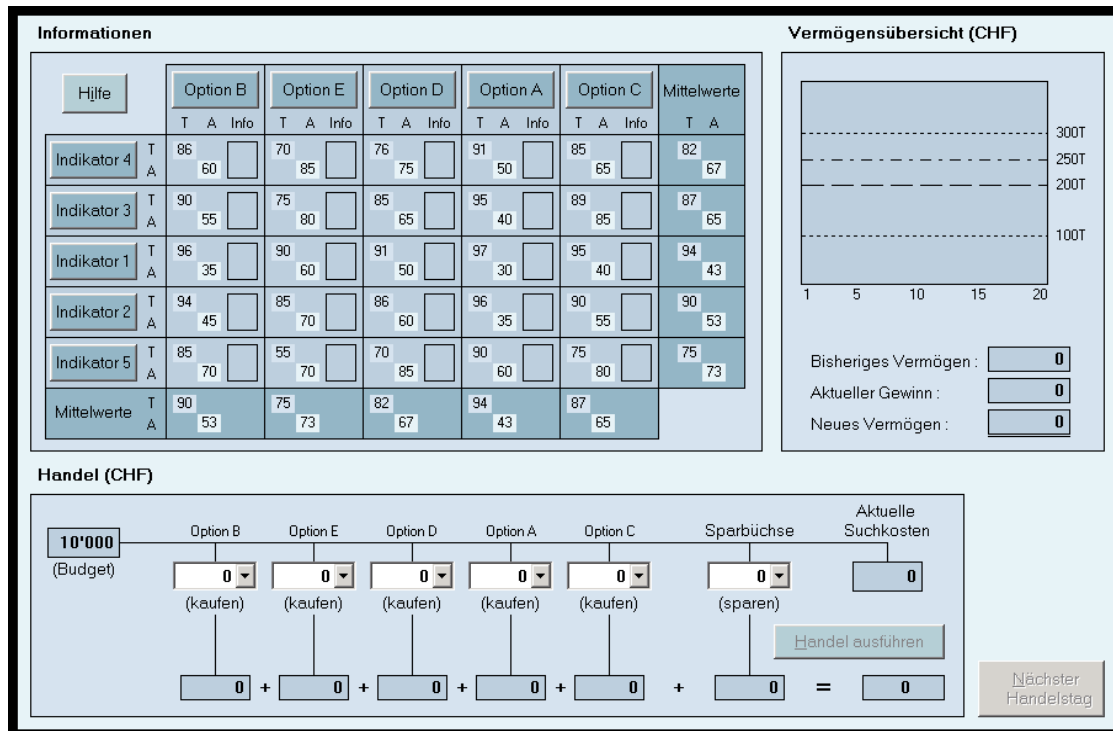


Abbildung 2.2: Bildschirmoberfläche von Experimentalbedingung B2 (Initialzustand)

In beiden Experimentalbedingungen haben die Versuchspersonen je 20 Durchgänge (Handelstage) zu absolvieren. In jedem Durchgang legen wir ihnen die fünf alternativen Optionsscheine bzw. „Optionen“ A, B, C, D und E vor. In diese gilt es möglichst geschickt zu investieren, um über alle Durchgänge einen möglichst hohen Gewinn zu erzielen. Für jeden Durchgang steht hierzu ein Spielgeld von CHF 10'000 zur Verfügung. Die Kursverläufe der Optionsscheine sind gegenseitig unabhängig. Pro Durchgang wird jeder Optionsschein im Wert entweder steigen oder sinken, jedoch niemals konstant bleiben.

Mit Hilfe von Kurs-Indikatoren entscheiden die Versuchspersonen in jedem Durchgang neu, ob sie in einen oder mehrere der zur Auswahl stehenden Optionsscheine investieren wollen und ob sie einen Teil oder gar ihr gesamtes Spielgeld direkt in die Sparbüchse legen möchten. Die Suche nach Cue-Informationen ist mit einem monetären Ressourcen-Aufwand verbunden: Jeder Informationsbeschaffungsvorgang schlägt mit CHF 1'000 zu Buche. Dies gilt natürlich nur für B2; in B1 liegen ja alle Informationen bereits offen vor und der Einsatz einer monetären Ressource zur Informationssuche ist damit hinfällig. Für B2 gilt somit, dass die Entscheidungssituation in der Sprache des One-Reason Decision Making (ORDM) formuliert ist. Das Verhältnis zwischen den Kosten für einen einzelnen Informationsbeschaffungsvorgang und dem verfügbaren Budget beträgt 1:10 und hat sich für ORDM gemäss unserer Erfahrung als favorabel erwiesen (vgl. Läge, Hausmann & Christen, 2005b). Abgesehen von den unterschiedlichen Informationskosten sind die statistischen Umwelten der beiden Experimentalbedingungen absolut identisch aufgebaut.

Die Cues sind auf dem information-board in einer 5x5-Matrix (5 Optionsscheine x 5 Indikatoren) angeordnet (wobei die Kopf- und Fusszeilen bzw. -spalten hier nicht mitgezählt sind). Pro Opti-

Optionsschein stehen also insgesamt 5 verschiedene Indikatoren zur Informationsbeschaffung zur Verfügung – und zwar für jeden Optionsschein dieselben fünf. Die von den Cues gelieferten Informationen werden durch Pfeile dargestellt. Ein schräg nach rechts oben zeigender, grüner Pfeil steht für eine positive Vorhersage, also dafür, dass ein Optionsschein an Wert gewinnen wird. Ein schräg nach rechts unten zeigender, roter Pfeil steht für eine negative Vorhersage, d.h. dafür, dass ein Optionsschein seinen Wert verlieren wird. Sofern ein konsultierter Cue keine Aussage zur Wertentwicklung eines Optionsscheins machen kann, wird ein schwarzes Kreuz (bestehend aus zwei diagonalen Linien) angezeigt.

Jeder Indikator ist pro Optionsschein mit einer spezifischen Trefferrate (Validität) behaftet, die angibt, mit welcher Wahrscheinlichkeit er für diesen Optionsschein eine korrekte Vorhersage bezüglich der Wertentwicklung liefert. Zusätzlich hat jeder Indikator, wiederum pro Optionsschein, eine spezifische Anwendbarkeit (Diskriminationsrate). Die Anwendbarkeit ist die Wahrscheinlichkeit, mit der der Indikator für den betreffenden Optionsschein überhaupt erst eine Aussage zur Wertentwicklung liefert, unabhängig davon ob diese richtig oder falsch ist. Ein Indikator liefert also zu einem bestimmten Optionsschein, sofern er für diesen überhaupt diskriminiert, eine dichotomisierte probabilistische Aussage über dessen Wertentwicklung (Wertsteigerung vs. Wertverlust).

In sämtlichen Durchgängen der beiden Experimentalbedingungen geben wir die Validitäten und Diskriminationsraten der Cues explizit auf dem Information Board bekannt. Dabei werden Validität und Diskriminationsrate in jedem der 5x5 Felder des Information Boards jeweils diagonal versetzt angezeigt. Auf diese Weise können wir die beiden Cue-Gütemasse spaltenweise (pro Option) bzw. zeilenweise (pro Indikator) zur Anzeige bringen, wobei in jeder Spalte (bzw. Zeile) entweder nur Validitätsangaben oder nur Diskriminationsraten untergebracht sind. Dadurch erhöhen wir die Übersichtlichkeit, so dass die Validitäten und Diskriminationsraten der Cues indikatorweise bzw. optionenweise schneller miteinander verglichen werden können. In den Zeilen- und Spalten-Köpfen der Tabelle ist in abgekürzter Schreibweise vermerkt, ob eine Zeile (bzw. Spalte) Validitätsangaben oder Diskriminationsraten enthält. Zeilen und Spalten mit Validitätsangaben sind mit dem Buchstaben "T" (für "Trefferrate") gekennzeichnet, Zeilen und Spalten mit Diskriminationsraten mit "A" (für "Anwendbarkeit"). Auch in den Instruktionen haben wir die beiden Fachbegriffe Validität und Diskriminationsrate durch die Begriffe Trefferrate und Anwendbarkeit ersetzt. Wie wir aus eigenen, bereits früher durchgeführten Experimenten wissen, sparen wir so einigen Erklärungsaufwand: Die Versuchspersonen verstehen dank dieser Massnahme rasch, welche Bedeutungen die beiden Gütemasse haben und wie sie diese Angaben zu ihrem Vorteil für das Börsenspiel nutzen können.

Die Validitäten und die Diskriminationsraten der Cues sind für das gesamte Experiment zeilen- und spaltenweise festgelegt und in beiden Experimentalbedingungen identisch, jedoch variiert von Durchgang zu Durchgang die Reihenfolge der Indikatoren sowie der Optionsscheine. Detaillierte Angaben zu den Cue-Validitäten und Diskriminationsraten können dem Kapitel "Technische Spezifikationen und Datenbasis" entnommen werden.

Abbildung 2.3 zeigt die in beiden Experimentalbedingungen gleichermassen verwendete Cue-Matrix. Die Matrix ist zeilenweise (bzw. spaltenweise) absteigend nach den mittleren Validitäten der Indikatoren angeordnet. Auf diese Weise ist die Gegenläufigkeit der Diskriminationsraten in Bezug auf die Validitäten gut zu erkennen.

Informationen																				
Hilfe			Option A			Option B			Option C			Option D			Option E			Mittelwerte		
			T	A	Info	T	A	Info	T	A	Info	T	A	Info	T	A	Info	T	A	
Indikator 1	T		97			96			95			91			90			94		
	A			30			35			40			50			60			43	
Indikator 2	T		96			94			90			86			85			90		
	A			35			45			55			60			70			53	
Indikator 3	T		95			90			89			85			75			87		
	A			40			55			85			65			80			65	
Indikator 4	T		91			86			85			76			70			82		
	A			50			60			65			75			85			67	
Indikator 5	T		90			85			75			70			55			75		
	A			60			70			80			85			70			73	
Mittelwerte	T		94			90			87			82			75					
	A			43			53			65			67			73				

Abbildung 2.3: In B1 und B2 verwendete Cue-Matrix (hier absteigend sortiert nach den Validitäten)

Liebe Versuchsteilnehmerin, lieber Versuchsteilnehmer

Herzlich willkommen bei unserem Experiment und vielen Dank für die Bereitschaft zur Teilnahme!

Die Forschungsfragestellung die hinter diesem Experiment steht, werden wir ganz am Ende erläutern.
Was Dich konkret erwartet und was Du zu tun hast, steht auf den folgenden Seiten.

[Zurück](#) [Start](#)

Abbildung 2.4: Kurzer Begrüssungstext zu Beginn des Experiments

2.3 Experimentalablauf

2.3.1 Überblick: Der Ablauf im Groben

Zu Beginn des Experiments erfassen wir mittels einer einfachen Eingabemaske den Namen der Versuchsperson sowie des Versuchsleiters / der Versuchsleiterin. Nach der Einblendung eines kurzen Begrüssungstexts (Abbildung 2.4) präsentieren wir die allgemeine Spielanleitung (Abbildung 2.5), in der wir der Versuchsperson erklären, worum es in dem Experiment geht.

Spielanleitung zum Börsenspiel

um die gesamte Instruktion zu lesen bitte am rechten Rand nach unten scrollen! >>

Aufgabe und Ziel

Du bist Teilnehmerin/Teilnehmer eines **vereinfachten Börsenspiels** und handelst mit **5 Optionsscheinen** (= Optionen A, B, C, D, E) **über insgesamt 40 Runden (= Handelstage)**. In jeder Runde steht Dir neues Geld im Umfang von **CHF 10'000** zur Verfügung (= **Budget**), das Du jeweils ganz auf eine Option setzen kannst oder aber auf mehrere Optionen gleichzeitig verteilen kannst. Nach Ablauf eines Handelstages siehst Du jeweils auf Deinem **Gewinnkontostand (= Vermögen)**, welche Option wie viel Gewinn abgeworfen hat. Dieses Börsenspiel ist so angelegt, dass nur **rund ein Drittel** aller Optionen über das gesamte Spiel hinweg an Wert **gewinnen** werden. Ist dies der Fall, bekommst Du den 3-fachen Betrag Deines Einsatzes auf Dein Gewinnkonto ausbezahlt. In zwei Drittel aller Fälle gewinnt die Option **nicht** und der von Dir eingesetzte Betrag ist verloren.

Dein Ziel besteht darin, im Laufe des Spiels einen möglichst hohen Gewinn zu erreichen.

Indikatoren, Trefferquote und Anwendbarkeit

5 Indikatoren (= Informationsquellen) geben Dir einen **Hinweis**, ob eine Option **Erfolg versprechend (steigender Pfeil)** oder **nicht Erfolg versprechend (sinkender Pfeil)** ist.

Die Verlässlichkeit dieser Informationsquellen (= Mittelwerte) und die der Einzelwerte wird durch die **Trefferquote** ausgedrückt. Die **Trefferquote** (eine Zahl zwischen 50% und 100%) gibt an, mit welcher Wahrscheinlichkeit von einem **Informationswert** auf eine korrekte Voraussage geschlossen werden kann. Beträgt die Wahrscheinlichkeit 100%, gewinnt die entsprechende Option mit Sicherheit. Bei einer Trefferquote von 50% könnte man genauso gut auch raten.

Nun kann es aber auch sein, dass ein Indikator einmal **keinen Hinweis** (= keine Information) liefert, d. h. es bleibt völlig im Dunkeln, ob die Option gewinnen oder verlieren wird. Diese Rate wird mit der **Anwendbarkeit** ausgedrückt. Die **Anwendbarkeit** (eine Zahl zwischen 0% und 100%) gibt an, mit welcher Wahrscheinlichkeit ein Indikator bzw. ein Einzelwert überhaupt eine Information liefert (d. h. ein sinkender oder fallender Pfeil).

Nun folgt ein Probedurchgang mit 3 Handelstagen!

Abbildung 2.5: Allgemeine Anleitung zum Börsenspiel

Anschliessend starten wir mit der ersten Experimentalbedingung, wobei die Reihenfolge der beiden Bedingungen über alle Versuchspersonen ausbalanciert ist, indem wir abwechselnd mit Experimentalbedingung 1 (B1) oder Experimentalbedingung 2 (B2) beginnen. In der folgenden Beschreibung des Experimentalablaufs wird zuerst B1 und anschliessend B2 durchgeführt.

Bevor die Versuchsperson mit B1 beginnen kann, muss sie einen 3 Durchgänge umfassenden Probelauf für diese Bedingung absolvieren. In diesem wird sie mit dem Aufbau der Bildschirmoberfläche, dem Ablauf der Versuchsdurchgänge und der Aufgabenstellung vertraut gemacht. Nach Abschluss des Probelaufs wird die spezifische Anleitung für B1 angezeigt (Abbildung 2.6). Anschliessend folgen die 20 Versuchsdurchgänge von B1. Abbildung 2.7 zeigt exemplarisch die Bildschirmoberfläche von B1 während eines Durchgangs.

Hinweis zur Phase mit vorgegebener Information

Der nun folgende Teil des Börsenspiels besteht aus 20 Runden.
Hier liegen alle **Informationen** zu den Optionen **bereits aufgedeckt** vor.

Dein Ziel besteht darin, mit dem jeweiligen Budget des Börsenspiels einen möglichst hohen Gewinn zu erreichen!

Du nimmst mit einem (weiteren) Los an einer **Verlosung von 3 x sFr. 100.- in bar** teil, wenn Du in dieser Phase **mind. CHF 300'000** erreichst.

Nun folgt der Hauptdurchgang!

Abbildung 2.6: Spezifische Spielanleitung zu Experimentalbedingung B1

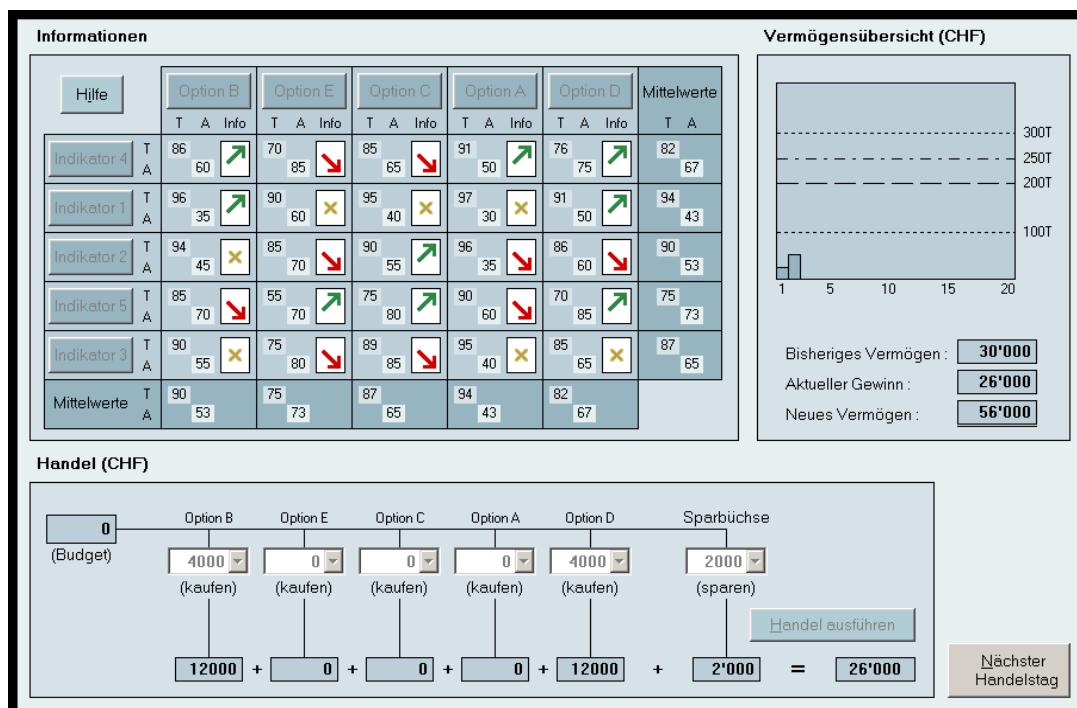


Abbildung 2.7: Bildschirmoberfläche von B1 (während eines Versuchsdurchgangs)

Danach geht es ohne zeitliche Verzögerung direkt mit B2 weiter: Auch hier führen wir mit der Versuchsperson zuerst einen Probelauf mit 3 Durchgängen durch und präsentieren ihr anschliessend die für B2 spezifische Spielanleitung (Abbildung 2.8). Anschliessend absolviert sie die 20

Versuchsdurchgänge von B2. Abbildung 2.9 zeigt exemplarisch die Bildschirmoberfläche von B2 während eines Versuchsdurchgangs.

Hinweis zur Phase der aktiven Informationssuche

Der nun folgende Teil des Börsenspiels besteht aus 20 Runden.

Hier müssen **Informationen** zu den einzelnen Optionen **aktiv gesucht werden**.

Das zeilen- oder spaltenweise Aufdecken von Informationen ist mit **Suchkosten von CHF 1'000** (pro Klick) verbunden.

Dein Ziel besteht darin, mit dem jeweiligen Budget des Börsenspiels einen möglichst hohen Gewinn zu erreichen!

Du nimmst mit einem (weiteren) Los an einer **Verlosung von 3 x sFr. 100 in bar** teil, wenn Du in dieser Phase **mind. CHF 300'000** erreichst.

Zielerreichung mit unterschiedlichen Strategien

In der **Phase der aktiven Informationssuche** wirst Du diese Limite mit folgenden Strategien bestimmt **nicht** erreichen:

1. Wenn Du dauernd sämtliche Informationen aufdeckst, bleiben Dir nur jeweils CHF 5'000 für das Setzen auf die Optionen. Da die durchschnittliche Trefferquote des besten Indikators unter 100% liegt, erreichst Du so bei 20 Runden einen Maximalgewinn von unter CHF 300'000 ($20 \times <15'000$).
2. Wenn Du gar keine Informationen aufdeckst, kannst Du zwar jeweils CHF 10'000 für das Setzen auf Optionen verwenden. Da die Gesamt-Trefferquote über alle Optionen und Runden aber nur ein Drittel beträgt, wird sich der Einsatz nur entsprechend zu 33% verdreifachen; dies ergibt einen Maximalgewinn von CHF 200'000 ($20 \times 1'000$).
3. Wenn Du den Handel ganz aussetzt und das gesamte Budget ausschliesslich in die Sparbüchse legt, kommen so am Ende nur CHF 200'000 ($20 \times 1'000$) zusammen.

Falls Du also an der Verlosung teilnehmen willst, musst Du eine andere, optimalere Strategie finden!

Nun folgt der Hauptdurchgang!

Abbildung 2.8: Spezifische Spielanleitung zu Experimentalbedingung B2

Informationen

	Option D			Option B			Option A			Option E			Option C			Mittelwerte	
	T	A	Info	T	A	Info	T	A	Info	T	A	Info	T	A	Info	T	A
Indikator 5	70	85		85	70		90	60		55	70		75	80		75	73
Indikator 2	86	60		94	45		96	35		85	70		90	55		90	53
Indikator 4	76	75		86	60		91	50		70	85		85	65		82	67
Indikator 3	85	65		90	55		95	40		75	80		89	85		87	65
Indikator 1	91	50		96	35		97	30		90	60		95	40		94	43
Mittelwerte	82	67		90	53		94	43		75	73		87	65			

Vermögensübersicht (CHF)

Bisheriges Vermögen: **65'000**
 Aktueller Gewinn: **24'000**
 Neues Vermögen: **89'000**

Handel (CHF)

Budget	Option D	Option B	Option A	Option E	Option C	Sparbüchse	Aktuelle Suchkosten
0	0	8000	0	0	0	0	2'000
(kaufen)	(kaufen)	(kaufen)	(kaufen)	(kaufen)	(kaufen)	(sparen)	

0 + 24000 + 0 + 0 + 0 + 0 = 24'000

Handel ausführen

Nächster Handelstag

Abbildung 2.9: Bildschirmoberfläche von B2 (während eines Versuchsdurchgangs)

Ein Durchgang läuft in beiden Experimentalbedingungen in den drei Phasen Informationssuche, Entscheiden und Feedback ab. Dementsprechend sind die Bildschirmoberflächen beider Bedingungen unterteilt in die drei Bereiche „Informationen“, „Handel“ und „Vermögensübersicht“. Die Oberfläche von B2 ist grundsätzlich identisch mit derjenigen von B1, unterscheidet sich von dieser jedoch darin, dass die Informationen hier nicht mehr offen vorliegen, sondern aktiv gesucht werden müssen. In den Zeilen- bzw. Spaltenköpfen der Daten-Matrix (im Bereich "Informationen") befinden sich deshalb an Stelle einfacher Labels Schaltflächen für den gezielten Informationsabruf.

Im Anschluss an den letzten Durchgang der zweiten Experimentalbedingung erheben wir auf der nächsten Bildschirmseite einige persönliche Angaben. Die Erhebung dieser Daten erfolgt mittels einer einfachen Eingabemaske (Abbildung 2.10). Die Daten werden direkt durch die Versuchsperson eingetragen. Sobald sie damit fertig ist, wird eine weitere Bildschirmseite angezeigt, auf der wir uns für die Teilnahme am Experiment bedanken (Abbildung 2.11). Auf der letzten Bildschirmseite schliesslich informieren wir die Versuchsperson über den von ihr im gesamten Experiment (d.h. in beiden Experimentalbedingungen zusammen) erreichten Spielstand (Abbildung 2.12).

Adresse & demographische Daten

Vorname: * Beruf (bzw. Studienfach / Semester): *

Nachname: * Geburts-Datum: * Geschlecht: *

Strasse + Nr. * Telefon-Nummer:

PLZ + Ort: * Email-Adresse:

Mit einem Stern (*) markierte Felder erfordern zwingend eine Eingabe

Teilnahme an zukünftigen Experimenten

☒ Ich bin bereit, mich für ein (allfälliges) Nachfolge-Experiment zur Verfügung zu stellen.

☐ Ich möchte in Zukunft nicht mehr an Experimenten an eurer Abteilung teilnehmen.

Grund (Angabe optional):

Kommentare

Mir ist bei eurem Experiment folgendes aufgefallen: / Vorschläge, Anmerkungen, etc:

Abbildung 2.10: Eingabemaske zur Erhebung einiger persönlicher Angaben



Abbildung 2.11: Bedankung für die Teilnahme am Experiment



Abbildung 2.12: Kurzer Begrüssungstext zu Beginn des Experiments

2.3.2 Detaillierte Betrachtung eines Versuchsdurchgangs

Informationssuche (Bereich „Informationen“). Im Bereich „Informationen“ wird also das Information Board dargestellt, über das die Versuchsperson entscheidungsrelevante Informationen beschaffen kann, wobei diese in B1 bereits offen vorliegen, in B2 hingegen aktiv und unter Aufwendung einer monetären Ressource gesucht werden müssen. Die Cue-Informationen können dabei zwar in beliebiger Reihenfolge, jedoch immer nur optionenweise oder indikatorenweise in Gruppen aufgedeckt werden. Das bedeutet, dass mit einem einzigen Mausklick jeweils immer alle 5 Informationsfelder einer Matrixzeile bzw. -spalte aufgedeckt werden. Mit jedem einzelnen Informationssuchschritt erhält die Versuchsperson somit entweder für eine ausgewählte Option die Informationen aller 5 Indikatoren (= optionenweise Informationssuche) oder für alle 5 Optionen jeweils die Informationen des ausgewählten Indikators (= indikatorenweise Informationssuche). Einmal aufgedeckte Informationen bleiben bis ans Ende des aktuellen Durchgangs auf dem Information Board sichtbar.

Nach einem Informationsabruf wird die entsprechende Schaltfläche deaktiviert, so dass sie im aktuellen Durchgang nicht erneut gedrückt werden kann. In jedem Durchgang kann eine Versuchsperson sämtliche Informationen bereits mit 5 Mausklicks (Suchkosten: 5 x CHF 1'000 = Total CHF 5'000) abrufen, indem sie bspw. fünfmal optionenweise vorgeht. Nicht gedrückte Schaltflächen (in o.g. Beispiel also die 5 Schaltflächen der Indikatoren) werden jedoch nicht deaktiviert. So ist es theoretisch durchaus möglich, dass eine Versuchsperson auch diese Schaltflächen noch drückt – was allerdings keinen Sinn macht, da sie dadurch keine zusätzlichen Informationen erhält. Für die Informationssuche können somit im Maximum CHF 10'000 pro Durchgang aufgewendet werden. Da die Informationssuchkosten vom Rundengeld abgezogen werden, stünde hernach für eine Investition in die Optionsscheine bzw. einen Transfer in die Sparbüchse keinerlei Geld mehr zur Verfügung. Es besteht jedoch kein Zwang, Informationen zu kaufen. So ist es auch erlaubt, ohne Suche nach Informationen in die verschiedenen Optionsscheine zu investieren (bzw. einen Betrag in die Sparbüchse zu transferieren). Die minimalen Informationssuchkosten betragen folglich CHF 0.

Im Erfolgsfall, d.h. bei steigendem Kursverlauf eines ausgewählten Optionsscheins, resultiert für diesen eine Gewinnausschüttung in dreifacher Höhe der getätigten Investition. Im Missergebnis resultiert für den betreffenden Optionsschein kein Gewinn und der in ihn investierte Einsatz ist verloren. Die durch die Investition in erfolgreiche Optionsscheine erzielten Gewinne werden sodann dem Gewinnkonto gutgeschrieben; ebenso ein allfällig in die Sparbüchse transferierter Betrag.

Entscheidungsphase (Bereich „Handel“). Nach der Phase der Informationsbeschaffung (= Such- und Stopp-Phase) treffen die Versuchspersonen im Bereich „Handel“ ihre Investitionsentscheidungen. Technisch gesehen ist das Ende der Informationsbeschaffungsphase durch die erste Aktivität der Versuchsperson im Bereich "Handel" festgelegt: Nach dem erstmaligen Setzen eines Betrages auf eine der zur Auswahl stehenden Alternativen werden die der Informationssuche dienenden Schaltknöpfe deaktiviert und es können keine Informationen mehr gesucht werden (auch wenn die Höhe des Betrags später ggf. wieder auf Null zurückgesetzt wird).

Die Versuchspersonen müssen in jedem Durchgang neu entscheiden, in welchen bzw. welche der insgesamt 5 zur Auswahl stehenden Optionsscheine sie wie viel Spielgeld investieren und welchen Betrag sie ggf. direkt in die Sparbüchse legen wollen. Zu Beginn eines jeden Durchgangs steht ihnen jeweils ein Spielgeld (Budget) von CHF 10'000 zur Verfügung. Was davon nach Abzug der aktuellen Informationssuchkosten noch übrig ist, muss restlos in die 6 Alternativen (5 Optionsscheine plus Sparbüchse) investiert werden. Das Spielgeld kann dabei in beliebiger Proportion auf beliebig viele (maximal alle 5, minimal 0) Optionsscheine und Sparbüchse verteilt werden. Im Maximum, d.h. wenn keine Informationssuche stattgefunden hat, kann das gesamte Budget von CHF 10'000 in eine einzige Alternative investiert werden kann. Falls eine Versuchsperson keinen der Optionsscheine positiv (d.h. als im Wert steigend) beurteilt, hat sie die Möglichkeit, den gesamten Betrag risikolos in die Sparbüchse zu legen. Investitionen in Optionsscheine sind risikobehaftet, bieten dafür aber die Möglichkeit, im Erfolgsfall einen Gewinn zu erwirtschaften. Investitionen, oder besser Transfers, in die Sparbüchse können hingegen ohne Risiko vorgenommen werden, bieten dafür allerdings keine Gewinnmöglichkeit.

Setzt eine Versuchsperson auf einen Optionsschein, der seinen Wert verliert, so verliert sie den vollen in diesen Optionsschein investierten Betrag. Investiert sie hingegen in einen Optionsschein, der im Wert steigt, so wird ihr ein Gewinn in dreifacher Höhe des investierten Betrags auf das Gewinnkonto gutgeschrieben. Ein direkt in die Sparbüchse transferierter Betrag wird ebenfalls (allerdings in unveränderter Höhe) dem Gewinnkonto gutgeschrieben.

In beiden Experimentalbedingungen geben wir jeweils denselben expliziten Zielzustand vor: Wir weisen die Versuchspersonen darauf hin, dass sie automatisch an der Verlosung von echtem Geld (3 x CHF 100) teilnehmen, sofern sie in mindestens einer Bedingung einen Gesamtgewinn von mindestens CHF 300'000 erzielen. Falls sie den vorgegebenen Zielzustand sowohl in B1 als auch in B2 erreichen oder überschreiten, nehmen sie mit zwei Losen an der Verlosung teil und haben damit eine doppelt so hohe Gewinnchance. Um einen möglichst hohen Spielgewinn (bzw. einen Spielgewinn von mindestens CHF 300'000) zu erreichen, sind die Versuchspersonen also gefordert, in jedem Durchgang möglichst viele gute Investitionsentscheidungen zu treffen.

Die Basisraten bezüglich der tatsächlichen Wertentwicklung der Optionsscheine teilen wir den Versuchspersonen ebenfalls mit: So wissen sie, dass die Wahrscheinlichkeit für eine positive Entwicklung (d.h. für die Wertsteigerung eines Optionsscheins) ein Drittel und für eine negative Entwicklung (Wertverlust) zwei Drittel beträgt. In beiden Experimentalbedingungen beginnen die Versuchspersonen jeweils in Runde 1 mit einem Gewinnkontostand von CHF 0.

Sobald eine Versuchsperson den insgesamt investierbaren Betrag auf die verschiedenen Alternativen aufgeteilt hat, muss sie noch die Schaltfläche "Handel ausführen" drücken, um die gewählte Aufteilung zu bestätigen. Solange sie diese Schaltfläche nicht gedrückt hat, kann sie die Aufteilung weiterhin verändern, danach jedoch nicht mehr. Sofort nach der Durchführung des Handels wird die Schaltfläche automatisch deaktiviert und die Versuchsperson erfährt das Ergebnis ihrer Investitionsentscheidung.

Feedback (Bereich „Vermögensübersicht“). Im Abschnitt "Vermögensübersicht" wird nun der über alle Optionen (inkl. Sparbüchse) summierte Gesamtgewinn des aktuellen Durchgangs angezeigt und die dargestellten Informationen (Neues Vermögen, grafische Anzeige des Vermögensverlaufes, etc.) werden auf den neusten Stand gebracht. Zudem wird die bisher deaktivierte Schaltfläche "Nächster Handelstag" aktiviert.

Für den nächsten Handelstag (d.h. den nächsten Durchgang) erhält die Versuchsperson erneut ein Budget von CHF 10'000. Der nächste Durchgang beginnt, sobald die Versuchsperson die Schaltfläche "Nächster Handelstag" drückt.

2.4 Technische Spezifikationen

Validitäten (V) und Diskriminationsraten (D). Die Validitäten und Diskriminationsraten sind für sämtliche Versuchspersonen und Durchgänge identisch, um die Vergleichbarkeit der beiden Experimentalbedingungen gewährleisten zu können. Die Cue-Matrix (d.h. die Matrix von Indikatoren x Optionen) ist wie folgt definiert: Zuerst ordnen wir die Indikatoren zeilenweise und die Optionen spaltenweise in aufsteigender Reihenfolge nach ihren Bezeichnungen (Nummern bzw. Buchstaben) an. Danach vergeben wir die Validitäten in absteigender Reihenfolge, sowohl pro Spalte als auch pro Zeile, so dass gleichzeitig auch die resultierenden mittleren Validitäten (der Indikatoren und der Optionen) bereits strikt absteigend angeordnet sind. Nun vergeben wir noch die Diskriminationsraten – sowohl pro Spalte als auch pro Zeile – in aufsteigender Reihenfolge, so dass sie indikatorweise und optionenweise gegenläufig zu den Validitäten sind. Auch hier achten wir darauf, dass (bei bestehender Sortierung) sowohl für die Indikatoren als auch die Optionen strikt aufsteigende mittlere Diskriminationsraten resultieren (vgl. Abbildung 2.3).

Der Bereich möglicher Cue-Validitäten erstreckt sich von > 0.50 bis < 1.00 . Abgesehen von einer einzigen Ausnahme ($V = 0.55$) liegen in diesem Experiment alle Validitäten im Bereich von 0.70 bis 0.97. Mit im Mittel 0.87 fallen sie recht hoch aus, so dass sich die Versuchspersonen gut auf bereits einen einzigen Hinweis verlassen können.

Die Diskriminationsraten der Cues liegen im Bereich von 0.30 bis 0.85 und sind in 5%-Schritten abgestuft. Da pro Indikator bzw. pro Option keine zwei Cues dieselbe Diskriminationsrate haben, unterscheiden sich alle Cues eines Indikators bzw. einer Option bezüglich ihrer Diskriminationsrate jeweils paarweise um mindestens 5%.

Betrachtet man nur die Validitäten der Cues, gibt es diesbezüglich einen besten Cue. Dasselbe gilt bezüglich der Diskriminationsraten. Aufgrund der Gegenläufigkeit der beiden Gütemasse ist der Cue mit der höchsten Validität aber gleichzeitig derjenige mit der geringsten Diskriminationsrate – und vice versa. Möchte man Cues die bezüglich einer Kombination von V und D (z.B. nach der Success-Formel, vgl. Martignon & Hoffrage, 1999) einen hohen Wert erreichen, so bieten sich die drei „mittleren“ Indikatoren (2, 3 und 4) bzw. Optionen (B, C und D) für die Informationssuche an.

Indikatoren, Cues und Optionen. Die Indikatoren sind von 1 bis 5 durchnummeriert und entsprechend benannt (i.e. „Indikator 1“, „Indikator 2“, etc.). Die Reihenfolge der Indikatoren (Zeilen) und der Optionsscheine (Spalten) variiert von Durchgang zu Durchgang, um den Versuchspersonen im Verlauf des Experiments kein erfolgreiches Lernen eines rein positionsgesteuerten Verhaltenspatterns zu ermöglichen. Durch gleichzeitiges Vertauschen der Zeilen untereinander sowie der Spalten untereinander erreichen wir eine verblüffend starke optische Veränderung der Daten-Anordnung in der Matrix. Durch diese Operationen werden die mittleren Validitäten und Diskriminationsraten der Indikatoren und Optionen natürlich nicht tangiert. Hinter den Bezeichnungen der Indikatoren und der Optionen verbergen sich über das gesamte Experiment und bei allen Versuchspersonen dieselben Cues bzw. Cue-Daten. Mit anderen Worten: „Indikator 1“ hat mit 0.94 immer dieselbe mittlere Validität und ist damit immer der valideste Indikator.

Reihenfolge der Experimentalbedingungen und der Durchgänge. Die Abfolge der Experimentalbedingungen (zuerst B1 und dann B2 oder umgekehrt) ist über alle Probanden ausbalanciert und in einer Datenbank abgespeichert: Versuchspersonen mit ungeraden Nummern beginnen mit B1, solche mit geraden Nummern mit B2.

Die in den 20 Durchgängen (Aufgaben) angezeigten Cue-Informationen sind allesamt vordefiniert (gemäß den festgelegten Basisraten, Cue-Validitäten und -Diskriminationsraten) und ebenfalls in der Datenbank abgespeichert. Allen Versuchspersonen werden jeweils dieselben 20 Aufgaben dargeboten, jedoch in unterschiedlichen Reihenfolgen. Wir haben für beide Experimentalbedingungen vordefiniert, in welchen Reihfolgen die Versuchspersonen die je 20 Durchgänge durchlaufen werden. Pro Versuchsperson sind jeweils zwei unterschiedliche Reihenfolgen definiert: Eine für B1 und eine für B2. Wir verwenden also in beiden Bedingungen dieselben 20 Aufgaben, da dies von keiner der im Anschluss an einen Vorversuch befragten Versuchspersonen bemerkt worden war. Die Reihenfolgen sind für beide Experimentalbedingungen separat ausbalanciert.

Aufgaben-Patterns für die beiden Experimentalbedingungen. Obwohl es für die Versuchspersonen bei 25 Cues und in nur 20 Durchgängen praktisch unmöglich ist, irgendwelche Validitäten und Diskriminationsraten systematisch zu beobachten oder gar zu lernen, sind dennoch die tatsächlichen Werte aus dem Experiment bestmöglich an die angegebenen Werte approximiert. (Eine perfekte Übereinstimmung ist natürlich bei nur wenigen diskriminierenden Fällen nicht auf das Hundertstel erreichbar, sondern nur im „long Run“ möglich.)

Die drei Cue-Ausprägungen (steigend, fallend, nicht diskriminierend) verteilen wir gemäß dem folgenden Zuordnungsverfahren auf die 20 verschiedenen Aufgabenpatterns: Zunächst wird für jeden Optionsschein in den 20 Patterns basisratenkonform (durch Ziehen ohne Zurücklegen) ausgelost, ob sein Wert steigt oder fällt, so dass über die 20 Durchgänge für jeden Optionsschein ein Drittel positive und zwei Drittel negative Ausprägungen resultieren. Aufgrund der Beschränkung der Anzahl Durchgänge auf genau 20, ist der Ausprägungsverlauf der Optionsscheine jedoch nur näherungsweise basisratenkonform: So resultieren $7/20 = 35\%$ positive und $14/20 = 65\%$ negative Ausprägungen pro Optionsschein. Anschliessend wird in den 20 Aufga-

benpatterns für jeden Cue entsprechend seiner Diskriminationsrate ausgelost, ob er diskriminiert oder nicht. In einem dritten Schritt wird schliesslich jedem diskriminierenden Cue validitätskonform zugelost, ob er eine korrekte oder eine falsche Vorhersage macht. Die den Versuchspersonen mitgeteilten Validitäten können, wie oben bereits festgestellt, jedoch nicht streng eingehalten werden: Die tatsächlichen Cue-Validitäten fallen deswegen allesamt etwas tiefer aus als die angezeigten, da die Cues rein rechnerisch unmöglich bestimmte Validitäts-Ausprägungen annehmen können: Indikator 1 bspw. hat für Option A eine Diskriminationsrate von 30% und diskriminiert deshalb (für Option A) lediglich in 6 der 20 Durchgänge. Dies hat zur Folge, dass die Validität (von Indikator 1 bezüglich Option A) über die 20 Durchgänge gemittelt nur den n -fachen Wert von $1/6$ betragen kann (wobei $n=0,\dots,6$). So beträgt die Validität denn auch tatsächlich nur $.83$ ($5/6$) anstelle des angegebenen Wertes von $.97$. Entsprechendes gilt, wenn auch nicht so extrem, auch für die anderen Validitäts-Angaben auf dem Information Board (inkl. Mittelwerte).

Der Unterschied zwischen den Validitäts-Angaben und den tatsächlichen Validitäten hat für das Experiment keine gravierenden Folgen, da die 20 Durchgänge im Verständnis der Versuchspersonen lediglich einen Ausschnitt aus einem grösseren Ausprägungszeitraum abbilden. Die Validitätsangaben stellen unter dieser Annahme durchaus adäquate Beschreibungen des tatsächlichen Sachverhalts dar. So haben sich im Anschluss an das Experiment auch keine Versuchspersonen über die angezeigten Validitäten verwundert gezeigt oder haben jemals an der Korrektheit der Angaben gezweifelt.

Zeitbedarfsabschätzung und Anzahl Durchgänge pro Bedingung. Die Abschätzung des realen Zeitbedarfs pro Versuchsdurchgang ist wichtig für die Bestimmung der Anzahl insgesamt durchführbarer Durchgänge pro Versuchsperson: Das ganze Experiment soll nicht länger als maximal 60 Minuten dauern. Um für unsere statistischen Auswertungen eine ausreichende Menge an Daten zu erhalten, wollen wir, dass eine Versuchsperson pro Bedingung 20 Durchgänge absolviert.

Für Bedingung B1 (Informationen liegen offen vor) rechnen wir mit einem Zeitbedarf von maximal ca. 20 Sekunden pro Durchgang. Inklusive Probedurchgang (3 Minuten), Durchlesen der spezifischen Instruktion (2 Min.) und Absolvieren der 20 Durchgänge (20×20 Sek. = ca. 7 Min.) sind damit für B1 maximal ca. 12 Minuten zu veranschlagen.

Für B2 (Informationen müssen aktiv gesucht werden) rechnen wir mit einem Zeitbedarf von maximal 40 Sekunden pro Durchgang. Dementsprechend wird die Absolvierung von 20 Durchgängen ca. 13 Minuten in Anspruch nehmen. Zusammen mit der benötigten Zeit für den Probedurchgang (5 Min.) und das Durchlesen der spezifischen Instruktion (3 Min.) wird B2 somit ca. 21 Minuten dauern.

Inklusive Begrüssung und Einleitung (5 Min.), Durchlesen der allgemeinen Instruktion (4 Min.), Erhebung von Angaben zur Person (4 Min.) und Verabschiedung (4 Min.), rechnen wir für das gesamte Experiment mit einem Realzeitbedarf von ca. 50 Minuten. Wir können somit problemlos insgesamt 40 Durchgänge pro Versuchsperson durchführen und damit deutlich unter der erwähnten maximalen Dauer von 60 Minuten für das gesamte Experiment bleiben.

Erwartungswerte unterschiedlicher Investitionsstrategien. Bei blinder Investition (d.h. einer Investition ohne vorausgehende Informationssuche) in einen oder mehrere Optionsscheine ist nach 20 Durchgängen ein mittlerer Gewinn von CHF 200'000 zu erwarten: Würde eine Versuchsperson in allen 20 Durchgängen jeweils ihr gesamtes Budget von CHF 10'000 immer auf dieselbe Option (Basisrate = .33) setzen, würde ein mittlerer „Gewinn“ von $20 \times .33 \times \text{CHF } 30'000 = \text{CHF } 200'000$ resultieren. Für blinde Investitionen in jeweils mehrere Optionsscheine pro Durchgang resultiert dieselbe Gewinn-Erwartung. Eine Versuchsperson, die ebenfalls in keinem der Durchgänge nach Informationen sucht und jeweils ihr gesamtes Budget (CHF 10'000) in die Sparsbüchse legt, hat nach 20 Durchgängen einen (sicheren) „Gewinn“ von CHF 200'000 auf ihrem Konto.

Um an der Verlosung von 3 x CHF 100 (echtes Geld) teilnehmen zu können, muss der Spielgewinn nach 20 Versuchsdurchgängen mindestens CHF 300'000 betragen. Dieser Betrag ist so gewählt, dass er bei vernünftiger Beachtung der verfügbaren Cue-Informationen gut zu erreichen ist, d.h. kein extrem riskantes Anlageverhalten erfordert. Die Versuchspersonen müssen also eine sinnvollere Strategie anwenden als blind zu investieren, wenn sie an der Verlosung teilnehmen wollen: Sie müssen nach Cue-Informationen suchen.

Die Aussicht auf 100 Franken Gewinn bei guter Leistung hat sich in unseren bisherigen Experimenten als brauchbare Alternative zu leistungskontingenter Bezahlung erwiesen. Bei leistungskontingenter Bezahlung hätten wir für jede Versuchsperson durchschnittlich 8-9 Franken Bargeld ausschütten können, was hinsichtlich der Belohnung nur geringe Varianzen ermöglicht hätte. Eine Belohnung nur der besten Versuchspersonen würde die eingesetzten Strategien notwendigerweise in Richtung auf extremes Risikoverhalten hin verschieben, was die Interpretation der Resultate unsicher machen würde. Der gewählte Modus liegt zwischen diesen Extremen.

2.5 Diskussion

Mit dem hier vorgestellten Design wird eine Möglichkeit geschaffen, neben dem Mehralternativenfall auch wahlweise optionsweises und attributweises Suchen zu ermöglichen. Auf diese Weise können die Informationssuchstrategien der Versuchspersonen effizient gemessen werden. Wie sich im durchgeführten Experiment gezeigt hat (vgl. Läge, Christen & Hausmann (2007)), zeigen die Versuchspersonen auch intraindividuell sehr stabile Strategien. Die Probanden können also gut mit dem Design umgehen.

Ferner ermöglicht das Programm eine beliebige Variation der Validitäten und Diskriminationsraten. Im durchgeführten Experiment lag hier kein besonderer Fokus, so dass sie gegenläufig verteilt waren. Im Grundsatz bedeutet dies, dass es sowohl zeilenweise als auch spaltenweise jeweils eine Möglichkeit gibt, die Informationssuche auf maximale Validität, maximale Diskriminationsrate oder eine Kombination aus beiden (im Sinne von Usefulness oder Success) auszurichten. Wenn man möchte, dann lassen sich mit diesem Design also auch die Suchregeln im Sinne eines Building Block einer Heuristik auf individueller Ebene erforschen.

Auch die Stopppregel kann – wie in der Auswertung von Läge, Christen & Hausmann (2007) gemessen werden. Da es in jeder Zeile / Spalte immer mindestens einen diskriminierenden Cue gibt, ist One Reason Decision Making für die Versuchspersonen möglich. Sie können aber auch mehr Information suchen, sei es, um Informationsintegration vorzunehmen, sei es, um eine durch die erste Suche aufgestellte Hypothese im Sinne der Risikominimierung noch einmal abzusichern.

Schliesslich ist auch bei der Entscheidungsregel eine gewisse Varianzbreite möglich. Im Zweialternativenfall lautet bei Urteilsheuristiken wie Take the Best ja die Entscheidungsregel: „Folge der Richtung des Hinweises“, was eigentlich trivial ist. Deswegen liegen auch noch keine interessanten Resultate im Umgang mit der Entscheidungsregel als einem der wichtigen Building Blocks vor. Mit der hier offerierten Möglichkeit, das Budget beliebig zu splitten oder auch auf die sichere Seite zu bringen, sind die Quellen für mögliche Varianz bereitet. Die Auswertungen von Läge, Christen & Hausmann (2007) geben bereits interessante Anhaltspunkte, in welche Richtung die Forschung in Sachen Entscheidungsregel laufen könnte.

Damit bietet das hier vorgestellte Design die Möglichkeit, alle Building Blocks einer Entscheidungsheuristik auf komplexerer Ebene untersuchen zu können als im bislang üblichen Zweialternativenfall.

2.6 Literatur

- Huber, O., Wider, R. & Huber, O. W. (1997). Active information search and complete information presentation in naturalistic risky decision tasks. *Acta Psychologica*, 95, 15-29.
- Huber, O., & Macho, S. (2001a). Probabilistic set-up and the search for probability information in quasi-naturalistic decision tasks. *Risk Decision and Policy*, 6, 1-16.
- Läge, D. Christen, S., & Hausmann, D. (2007). *Der Einfluss der Informationssuche auf die Entscheidung*. AKZ-Forschungsbericht 04. Zürich: Angewandte Kognitionspsychologie.
- Läge, D., Hausmann, D., & Christen, S. (2005b). *Wie viel bezahlen für eine valide Information? Suchkosten als limitierender Faktor der Informationssuche*. AKZ-Forschungsbericht 07. Zürich: Angewandte Kognitionspsychologie.
- Martignon, L., Hoffrage, U. (1999). Why does one-reason decision making work? In: G. Gigerenzer, P.M. Todd and the ABC Research Group. *Simple Heuristics that Make Us Smart*. New York: Oxford University Press. 119-140.
- McDougal, Y. B. (1995). Decision making under risk: risk preference, monetary goals and information search. *Personality and Individual Differences*, 18, 771–782.
- Payne, J. W. (1976). Task complexity and contingent processing in decision making: An information search and protocol analysis. *Organizational Behavior and Human Performance*, 16, 366-387.

Kapitel 3

Der Einfluss der Informationssuche auf die Entscheidung

Im Alltag ist die vorgängige Suche nach Information eine zentrale Voraussetzung für die meisten Entscheidungen und wird deshalb in der theoretischen Modellierung vielfach als Phase des Entscheidungsprozesses betrachtet. In der Mehrzahl der klassischen Entscheidungsexperimente wird jedoch die aktive Informationssuche ausgeklammert und stattdessen sämtliche relevante Information vorgegeben. Vor diesem Hintergrund präsentieren wir ein Experiment, welches die Bedeutung der aktiven Informationssuche für die spätere Entscheidung aufzeigen soll: Führt – bei identischer Informationsstruktur im Hintergrund – Entscheiden auf der Basis vorgegebener Information (Bedingung 1) zu einem systematisch anderen Verhalten als bei einer aktiven Suche nach Information (Bedingung 2)?

In einem Börsenspiel können Personen einen Geldbetrag auf eine oder auf bis zu fünf Optionsscheine setzen. In Bedingung 1 bekommt jede Versuchsperson für die fünf Optionen sämtliche Ausprägungen zu insgesamt fünf probabilistischen Hinweis-Cues angezeigt, in Bedingung 2 muss sie (dieselbe) Information aktiv und sequentiell aufdecken. Dabei wird die Entscheidungsvarianz zwischen diesen beiden Bedingungen intrapersonell derart gross, als würde die Person gegen eine andere ausgetauscht!

Ferner wurden intrapersonell sehr stabile Muster der Informationssuche gefunden: Knapp die Hälfte der Probanden testet sukzessive Optionsschein für Optionsschein mit all seinen Informationen (eine Strategie, die dem Satisficing von Simon nahe kommt), während etwa gleich viele Personen das Informationsfeld attributweise absキャンen, indem sie sich einen Cue für alle Optionsscheine anschauen (je nach Stoppregel ist dies als One Reason Decision Making oder als Informationsintegration interpretierbar). Die Suchstrategien scheinen auch Einstellungen zum Risiko widerzuspiegeln: Personen, welche bevorzugt optionsweise suchen, sind eher risiko-vermeidend und verteilen einen Geldbetrag auf mehr Optionsscheine bzw. legen einen höheren Betrag in die sichere Sparbüchse. Personen, welche das Feld attributweise scannen, zeigen sich vergleichsweise risikosuchender.

3.1 Einleitung

Wie viel Information benötigen Sie, damit man Sie für eine möglicherweise sehr lukrative, aber in jedem Fall sehr riskante Geldanlage überzeugen kann? Reicht es Ihnen aus, wenn eine vertrauenswürdige Person mit „ich habe da einen guten Tipp erhalten“ kommt, oder wollen Sie doch eine möglichst vollständige Analyse aller verfügbaren Informationen über die Aussichten der betreffenden riskanten Geldanlage?

Letzteres erfordert natürlich das Zusammentragen dieser Informationen sowie deren sorgfältige Analyse. Doch dafür sollten Sie dann auch die bestmögliche Grundlage für Ihre Entscheidung haben. So ist diese Vorgehensweise denn auch das klassische Ideal des Entscheidens unter Risiko: Alle möglichen Ausgänge einer Entscheidung werden berücksichtigt und in der Wünschbarkeit der erwartbaren Konsequenzen unter Berücksichtigung deren Eintrittswahrscheinlichkeiten zu einem optimalen Gesamturteil integriert. Dass Menschen substantielle Schwierigkeiten haben, diese Norm eines nutzen-maximierenden *Homo oeconomicus* (Mill, 1848) in der Realität einzulösen, ist spätestens seit den epochemachenden Arbeiten von Tversky & Kahneman (1974) sowie durch das von Kahneman, Slovic & Tversky (1982) ausgelöste Forschungsprogramm der *Heuristics and Biases* bekannt. Das in diesem Programm beschriebene systematische Fehlverhalten lässt sich auf Populationsebene ausnützen und wurde deshalb unter dem Namen *Behavioral Finance* in den Wirtschaftswissenschaften gerne aufgenommen (vgl. Übersichten in Goldberg & Nitzsch, 2000; Kiehling, 2001; Kirchler, 1999; Shefrin, 2000; Unser, 1999).

Neben diesen Regelverletzungen, die systematisch auftreten, wenn man Personen in bestimmter Weise mit den Normen einer wahrnehmungstheoretisch anspruchsvollen Analyse konfrontiert, finden sich aus psychologischer Sicht aber auch noch einige ganz andere Ansätze, um mit der Frage guten Entscheidens umzugehen. Sie verleihen dem Menschen ein wenig von der Würde eines „rationalen Wesens“ zurück, die er durch die Befunde des Heuristics and Biases-Programm ganz offensichtlich verloren hatte. Dawes (1979) beispielsweise zeigte auf, dass man bei der rechnerisch schwierigen Integration probabilistischer Informationen auch ohne Angabe von Gewichten auskommen kann und mit dem einfachen Auszählen der Gründe pro und kontra eine vernünftige Annäherung an eine im Rechenaufwand immense Berücksichtigung des Bayes-Theorems erhält. Die nach ihm benannte *Dawes' Rule* (auch als *tallying* bezeichnet) ist denn ein einfaches Auszählverfahren der positiven wie negativen Argumente. Damit, so das implizite Argument, sollten Menschen ja wohl klar kommen.

Hinsichtlich der Norm bleibt Dawes jedoch dem Rational einer vollständigen Informationsverarbeitung verhaftet. Simon (1955, 1956, 1957) hat hingegen bereits früh darauf hingewiesen, dass vollständige Informationsverarbeitung in realen Umwelten oftmals weder möglich noch nötig ist. Ersteres exemplifiziert er an sequentiell vorliegenden Alternativen, bei denen man sich für oder gegen eine Option entscheiden muss, noch bevor man alle anderen gesehen hat. In diesen Fällen ist es gar nicht möglich, vollständige Information über alle Alternativen zu integrieren, um zu einer optimalen Entscheidung zu gelangen. Letzteren Aspekt (dass eine optimale Entscheidung oft gar nicht nötig ist) führt er am Beispiel der unterschiedlichen Ziele aus, die ein Organismus

zu realisieren versucht: Geht es etwa um die Befriedigung homöostatischer Bedürfnisse, so braucht eine Entscheidung nur zufriedenstellend zu sein, um dem Organismus auszureichen (Simon bezeichnet dies als „satisficing“).

Einen dritten, noch radikaleren Weg schlägt die Gruppe um Gigerenzer ein (z.B. Gigerenzer, Todd & the ABC Research Group, 1999). Sie führt Simons Gedanken der Angemessenheit einer Entscheidung an bestehende Ziele (und indirekt auch das Argument von Dawes hinsichtlich der Einfachheit der Mittel) konsequent fort, indem sie auf einfache Urteilsheuristiken setzt, mit denen genügend gute Entscheidungen in probabilistischen Umwelten möglich werden. Diese Heuristiken sind – ganz anders als die des *Heuristics and Biases*-Ansatzes – als psychologische Prozessmodelle spezifiziert, die auf Informationssuche, Abbruch der Suche und Durchführen einer Entscheidung auf der Basis der bis dahin gefundenen Informationen beruhen. Diese sogenannten *Building Blocks* (Suchregel, Stoppregel und Entscheidungsregel) einer Urteilsheuristik stellen das Entscheiden in ein ganz anderes Licht, als es in klassischen Informationsintegrationsmodellen der Fall ist: Entscheidend ist nicht allein die Verrechnung der vorliegenden Information, sondern vielmehr der Prozess, mit dem ein Organismus in seiner Umwelt an die relevante Information gelangt. Informationssuche wird also zur condition sine qua non der Entscheidungspsychologie.

Gleichzeitig wird in den von Gigerenzer vorgeschlagenen Urteilsheuristiken die Entscheidungssituation drastisch vereinfacht: Die Heuristiken stoppen nämlich, sobald eine einzige zwischen den Alternativen diskriminierende Information vorliegt. Dieses sogenannte *One-Reason Decision Making* (ORDM) wird am eindrucklichsten durch die *Take The Best-Heuristik* (TTB) beschrieben (hergeleitet von Gigerenzer, Hoffrage & Kleinbölting, 1991, auf der Basis der lexikographischen Entscheidungsregel von Fischhoff, 1975). Sie setzt sich aus den drei *Building Blocks* „Suche nach der validesten Information“, „Stoppe, sobald Du eine diskriminierende Information gefunden hast“ und „Entscheide Dich in Richtung dieser Information“ zusammen. Dass dieses Paradebeispiel von ORDM zu praktisch gleich guten Entscheidungen führt wie eine vollständige Informationsintegration durch ein lineares Regressionsmodell, konnte von Gigerenzer & Goldstein (1996) in einer Simulation am Beispiel des Zwei-Alternativenfalls der Entscheidung über die Einwohnerzahlen deutscher Grossstädte gezeigt werden.

Gigerenzer wäre – um zum Ausgangsbeispiel zurückzukehren – also durchaus für den einen guten Tipp zu haben, um darauf eine Finanzentscheidung aufzubauen. In der Tat haben Gigerenzer & Goldstein (1999) nicht nur einen Aufsatz „Betting on one good reason“ zur TTB-Heuristik verfasst, sondern Borges, Goldstein, Ortmann & Gigerenzer (1999) haben allein aufgrund der *Recognition Heuristic* Geld an der Börse gesetzt – und damit gewonnen. Sie sehen ORDM – vor allem in der Einkleidung der TTB-Heuristik – also für eine absolut valable Alternative zur aufwändigen Informationsintegration.

Die empirischen Evidenzen für den tatsächlichen Einsatz von TTB im Entscheiden sind allerdings bislang recht gering. Bröder gibt einen hervorragenden Überblick über die vorliegenden Experimente und kommt zu dem Schluss, dass TTB zwar zum menschlichen Verhaltensreper-

toire gehöre, in seiner Reinform im Experiment aber nur unter bestimmten Rahmenbedingungen (vornehmlich unter Informationskosten) gehäuft auftrete und sich dann zumeist nur bei einem gewissen Anteil der Probanden und zudem noch intraindividuell inkonsistent zeige (Bröder, 2005, S. 70f., S.81). Läge & Daub (2006) zeigten, dass ohne vorherige aktive Informationssuche (bei TTB ein essentieller Aspekt!) ohnehin keine Begeisterung dafür zu bestehen scheint, sich allein auf die valide Informationsquelle zu verlassen. Genauso wie Lee & Cummins (2004) fanden die Autoren intraindividuell sehr uneinheitliches Verhalten sowie eine erdrückende Mehrzahl der Probanden, die im Zweifelsfall die Informationsintegration – vornehmlich im Sinne der *Dawes' Rule* – einer Beachtung des validesten Cues vorzog (nur eine einzige von 58 Versuchspersonen wendete ORDM an). Dieses Resultat basierte jedoch auf einem Experiment, bei dem alle relevanten Informationen vorgegeben waren. Läge, Hausmann, Christen & Daub (2005) verwendeten dasselbe statistische Szenario in einem Experiment mit aktiver Informationssuche und fanden dort mehrheitlich ORDM (in 78% der Fälle).

Informationssuche scheint also ein kritischer Faktor zu sein, so dass es möglich erscheint, Entscheiden bei vorliegender Information und Entscheiden bei aktiver Informationssuche als zwei unterschiedliche Situationsklassen bewerten zu müssen. Das Konzept der Informationssuche wurde in der Entscheidungspsychologie bereits mehrfach diskutiert (z.B. Huber, Wider & Huber, 1997; sowie Huber & Macho, 2001; Roetheli (2002) erwägt sogar, im Anschluss an das Satisficing-Konzept Simons die eigentliche Rationalität im Entscheidungsprozess von der Informationsverarbeitung hin zur Informationsaufnahme und zum optimalen Zeitpunkt des Abbruchs der Informationssuche zu verlagern). Jedoch sind uns keine Experimente bekannt, die einen direkten Vergleich dieser beiden Bedingungen im Hinblick auf das resultierende Entscheidungsverhalten leisten.¹⁶ Im Folgenden wird deswegen ein Experiment präsentiert, welches genau dies zum Ziel hat: Es soll herausfinden, ob – bei identischer Informationsstruktur im Hintergrund – Entscheiden auf der Basis vorgegebener Information (Bedingung 1) zu systematisch anderem Verhalten führt als bei einer aktiven Suche nach Information (Bedingung 2) (Hauptfragestellung).

Im Hinblick auf die Entscheidungsregel wäre zu prüfen, ob Personen stärker im besten bzw. validesten Argument (im analogen Sinne zu TTB) oder in einer Integration der insgesamt vorliegenden Information (z.B. durch einfaches Abzählen der Argumente) den Weg zum Erfolg sehen (Fragestellung 2).

Neben der Hauptfragestellung und einer eher allgemeinen Betrachtung des Entscheidungsverhaltens könnte das Auffinden individueller Suchstrategien von Interesse sein. Zumindest erscheint es uns interessant zu prüfen, ob sich Personen im Falle der Informationssuche und der Stopppregel eher zu einem attributswise suchenden One-Reason Decision Making, zu einer sequentiell opti-

¹⁶ Die wenigen Experimente zur aktiven Informationssuche liegen entweder im Paradigma der Ökologischen Rationalität vor (vgl. den Überblick bei Bröder, 2005), beziehen sich auf streng sequentiell vorliegende Alternativen (Johnson, 1989), oder sie haben vornehmlich darauf fokussiert zu zeigen, dass Versuchspersonen nur einen Teil der potentiell vorhandenen Informationsquellen abfragen (z.B. Edwards & Slovic, 1965; Frey, 1981; Fried & Peterson, 1969; Huang, 2000; Kerstholt, 1992; Kirschenbaum, 1992; Sheluga, Jaccard & Jacoby, 1979; Sundström, 1987).

onsprüfenden Vorgehensweise im Sinne eines Satisficing oder zu einer Überprüfung des vollständigen Informationsraumes (beispielsweise mit Anwendung einer vereinfachten Integrationsregel wie der Dawes' Rule) hingezogen fühlen (Fragestellung 3).

Im Zusammenspiel zwischen Informationssuchprozess und Entscheidungsverhalten schliesslich erscheint es uns interessant, ob bestimmte intendierte Entscheidungsmuster ihrerseits bestimmte Informationssuchstrategien auslösen. Ein geeigneter Kandidat für einen derartigen Zusammenhang scheint uns die Risikobereitschaft im Sinne eines Persönlichkeitsmerkmals zu sein (vgl. Übersichten in Jungermann, Pfister & Fischer, 2005; Eisenführ & Weber, 2003; Huber, Beutter, Montoya & Huber, 2001). Untersuchungen von McDougal (1995) zeigten z.B., dass Risiko-averse Personen sich ein genaueres Bild über einzelne Optionen machen (in der Terminologie der geplanten Untersuchung also zur Informationsintegration tendieren), während Risikogeneigte Personen möglichst viele Optionen anhand einiger weniger Merkmale bewerten (also eher zum ORDM neigen sollten). Als Fragestellung 4 werden wir daher in die Untersuchung aufnehmen, ob riskanteres Verhalten (d.h. alles Geld auf eine Option zu setzen bzw. weniger Geld in eine sichere Sparbüchse zu retten) mit einer höheren Tendenz zu einer Informationssuchstrategie im Sinne des ORDM einhergeht.

3.2 Methoden

Am Experiment nahmen 35 Versuchspersonen (15 Frauen und 20 Männer) im Alter zwischen 20 und 39 Jahren ($M = 27.5$ Jahre) teil, zumeist Studierende der Philosophischen Fakultät der Universität Zürich. Sie lösten in zwei unterschiedlichen Durchgängen insgesamt 40 Investitionsaufgaben auf einem Information-Board (vgl. Abbildung 3.1 – 3.3). Pro Aufgabe stand jeweils ein Budget von (fiktiven) 10'000 Schweizer Franken zur Verfügung, mit dem bis zu fünf Optionsscheine gekauft werden konnten. Die Optionsscheine verdreifachten im positiven Ausgang ihren Wert, verfielen aber im negativen Fall vollständig. Den Versuchspersonen wurde erklärt, dass die Wahrscheinlichkeit für einen positiven Ausgang (Gewinn) ein Drittel beträgt, für einen negativen Ausgang (Verlust) zwei Drittel. Ziel war es, in jedem der beiden Durchgänge einen maximalen Betrag zu erwirtschaften.

Fünf verschiedene Optionsscheine standen in jeder Aufgabe zur Verfügung. Die Versuchsperson konnte ihr Budget auf einen dieser Optionsscheine setzen, in beliebiger Proportion auf mehrere Optionsscheine verteilen oder das Geld ganz oder teilweise in die Sparbüchse stecken. Wenn die Vp das Geld verteilt hatte, quittierte sie mit einem entsprechenden Button und sah unmittelbar das Ergebnis ihrer Investition: erfolgreiche Optionsscheine verdreifachten den Wert, Beträge in der Sparbüchse blieben unverändert, und das in erfolglose Optionsscheine investierte Geld verfiel. Der summierte Gesamtgewinn aus diesem Durchgang wurde angezeigt und auf dem information board oben rechts zum bisher erwirtschafteten Vermögen hinzuaddiert. Die Versuchsperson erhielt dann für die nächste Aufgabe ein neues Budget von 10'000 Franken. Bei einer völlig zufälligen Investitionsstrategie – wie auch bei einer Einlage in die Sparbüchse – wäre der Erwartungswert null gewesen, d.h. der erwartete Outcome wäre genau so hoch gewesen wie das total zur Verfügung gestellte Budget (200'000 Franken pro Durchgang mit total 20 Aufgaben). Um

der Versuchsperson zusätzlich zur verbalen Instruktion der Gewinnmaximierung einen Anreiz zu bieten, erhielt sie für jeden Durchgang, bei dem sie zumindest 300'000 Franken erwirtschaftete, einen Teilnahmechein für eine Auslosung von drei Prämien zu je 100 Schweizer Franken (diesmal echtes Geld). Dieser Betrag (300'000 Franken) war so gewählt, dass er mit vernünftiger Beachtung der verfügbaren Hinweisinformationen gut erreichbar war, also kein extrem riskantes Anlageverhalten erforderte.¹⁷

Um nicht raten zu müssen, standen den Versuchspersonen für jeden Optionsschein fünf Hinweis-Cues zur Verfügung (als „Indikatoren“ eingeführt), die etwas über die Wahrscheinlichkeit aussagen, dass sich ein bestimmter Optionsschein positiv oder negativ entwickeln wird. Zu jeder Cue-Information war entsprechend die ökologische Validität angegeben (der Versuchsperson als Trefferrate der vom Cue gemachten Vorhersage erklärt) sowie die Diskriminationsrate (als Anwendbarkeit eingeführt, also als Prozentsatz, in dem ein Cue überhaupt eine Vorhersage ermöglicht). Der Versuchsperson wurde vereinfachend auf dem Information Board dargestellt, ob der betreffende Indikator für einen positiven Ausgang spreche (gelber Pfeil nach oben), für einen negativen Ausgang (roter Pfeil nach unten), oder ob dieser Indikator in diesem Fall für den betreffenden Optionsschein nicht anwendbar sei bzw. keine Information liefere (schwarzes X). Die Validitäten der Cues fielen im Mittel mit .87 recht hoch aus. Die Versuchspersonen konnten sich mit einer hohen Trefferrate also bereits auf einen einzigen Hinweis verlassen.

Die Validitäten und Diskriminationsraten wurden zeilen- und spaltenweise für das gesamte Experiment festgelegt. Zeilen und Spalten variierten jedoch von Aufgabe zu Aufgabe in der Anordnung, um der Versuchsperson kein erfolgreiches Lernen eines rein positionsgesteuerten Verhaltenpatterns im Verlauf des Experiments zu ermöglichen. (Durch das gleichzeitige Vertauschen der Reihenfolge in den Zeilen und Spalten ändert sich die Anordnung in der Tat sehr stark.) In den Abbildungen 3.1 – 3.3 sind die Zeilen und Spalten in absteigender Reihenfolge der Validitäten und aufsteigender Reihenfolge der Diskriminationsraten sortiert. Auf diese Weise wird dem Betrachter klar, dass es hinsichtlich der Validität einen besten Cue gab und einen dazu gegenläufigen mit der höchsten Diskriminationsrate. Die drei Cues dazwischen boten sich an, wenn man sowohl eine gute Validität als auch eine gute Diskriminationsrate haben wollte (z.B. nach der Success-Formel, vgl. Martignon & Hoffrage, 1999).

Die drei Cueausprägungen (positiv, negativ, nicht diskriminierend) wurden gemäss der angegebenen Validitäten und Diskriminationsraten zufällig auf 20 verschiedene Aufgabenpattern verteilt, indem für jeden Optionsschein in den 20 Pattern zunächst ausgelost wurde, ob er positiv oder negativ ausgeht (ein Drittel positiv, zwei Drittel negativ), woraufhin dann mit der Wahrscheinlichkeit der Diskriminationsrate ausgelost wurde, ob dieser Cue für diesen Optionsschein

¹⁷ Die Aussicht auf 100 Franken Gewinn bei einer guten Leistung hat sich in unseren bisherigen Experimenten als valable Alternative zur leistungskontingenten Bezahlung erwiesen. In diesem Fall hätten wir für jede Versuchsperson durchschnittlich acht bis neun Franken ausschütten können, was hinsichtlich der Belohnung nur geringe Varianzen ermöglicht hätte. Eine Belohnung nur für die besten Versuchspersonen hätte die eingesetzten Strategien notwendigerweise in Richtung auf extremes Risikoverhalten hin verschieben können, was die Interpretation der Resultate unsicher macht.

in diesem Pattern diskriminierte oder nicht, bevor schliesslich in einem dritten Schritt mit der Wahrscheinlichkeit der Validität zugelost wurde, ob der diskriminierende Wert die richtige oder die falsche Vorhersage machte. Im Ergebnis wurden also die Angaben, die den Versuchspersonen gegenüber hinsichtlich Validität und Diskriminationsrate gemacht wurden, streng eingehalten.

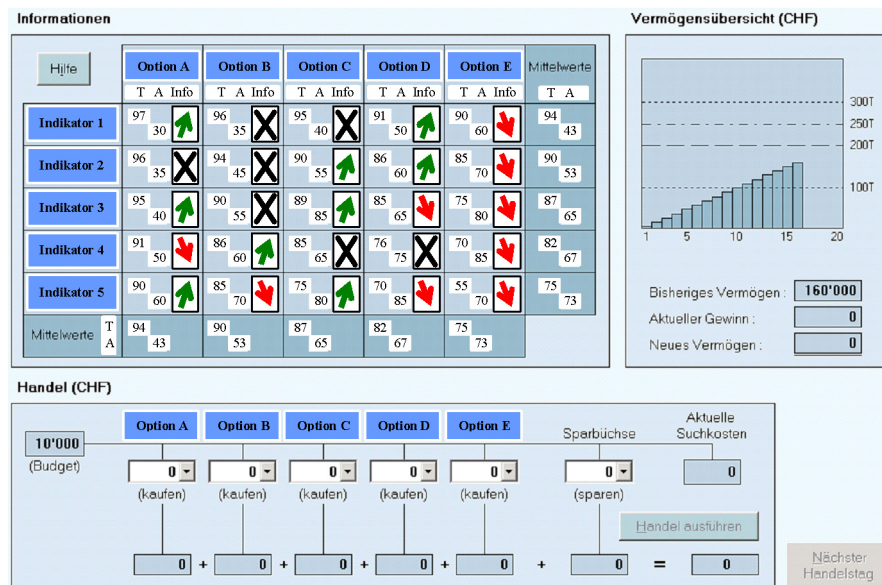


Abbildung 3.1: Informationboard der Bedingung 1 mit vollständiger Information (ohne Informationssuche)

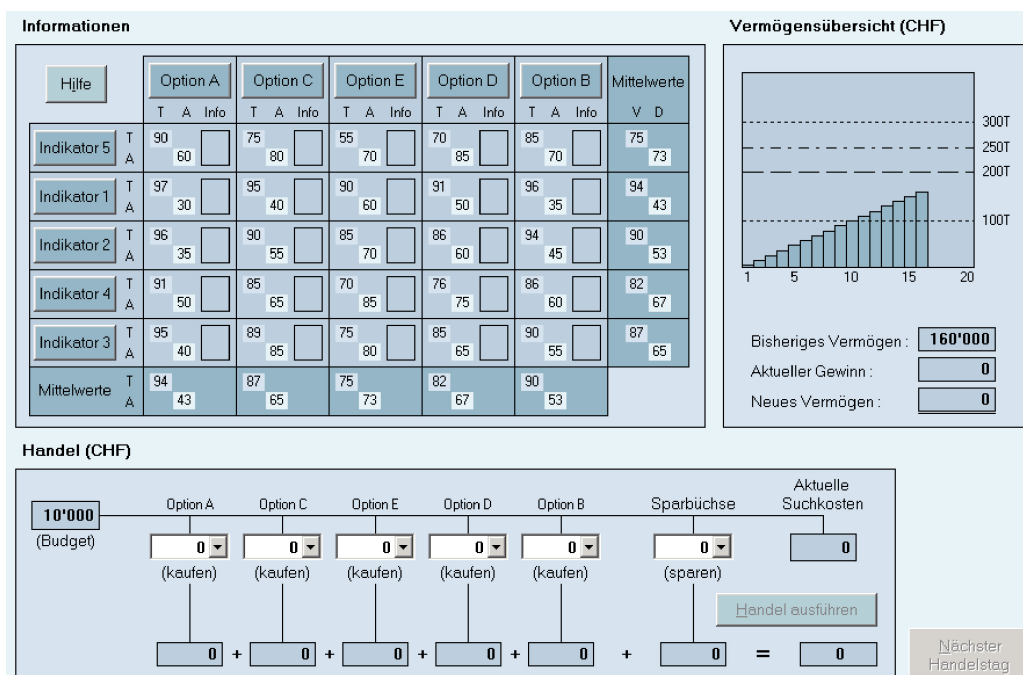


Abbildung 3.2: Leeres Information Board in Bedingung 2 (vor aktiver Informationssuche)

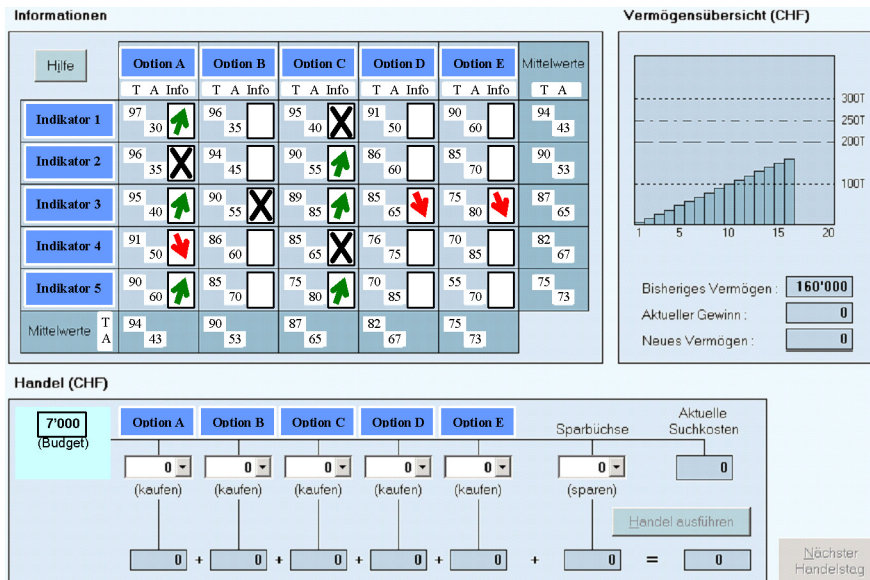


Abbildung 3.3: Information Board in der Bedingung 2 (nach aktiver Informationssuche und vor der Investitions-Entscheidung): Als erstes wurde der Indikator 3 mit einer mittleren Trefferquote (T) von .87 und einer mittleren Anwendbarkeit (A) von .65 aufgedeckt. Anschließend wurden die beiden Optionen A und C näher evaluiert. Der verbleibende Betrag von 7'000 Franken (Startkapital minus Suchkosten) kann nun auf eine oder mehrere der Optionen gesetzt oder (vollständig bzw. teilweise) in die Sparbüchse gelegt werden.

Es resultierten 20 Informationspatterns, die von jeder Versuchsperson zweimal bearbeitet wurden. In Bedingung 1 erhielt sie in jeder Aufgabe alle 25 Cue-Informationen direkt mit der Aufgabe in der 5x5-Matrix angezeigt (Abbildung 3.1). In Bedingung 2 sah sie zuerst ein leeres Information Board (Abbildung 3.2) und konnte (durch Anklicken des betreffenden Indikator- oder Optionsschein-Feldes) für jeweils 1'000 Franken¹⁸ alle fünf Informationen zu einem Cue oder alle fünf Informationen zu einem Optionsschein kaufen. In Abbildung 3.3 beispielsweise hat die Versuchsperson zunächst den Indikator 3 gekauft. Er erbrachte zwei positive Informationen (Optionsschein A und C). Für diese beiden Optionsscheine hat die Person sich dann (für weitere zwei mal 1'000 Franken) alle Cue-Informationen angesehen. Der jeweils für die Informationssuche investierte Betrag wurde vom Budget (10'000) abgezogen und konnte entsprechend nicht mehr angelegt werden. Im Beispiel von Abbildung 3.3 standen der Versuchsperson für die Anlage somit noch 7'000 Franken zur Verfügung.

Die Reihenfolge der Aufgabenblöcke der beiden Bedingungen wurde über die Versuchspersonen ausbalanciert und die 20 Aufgaben pro Block jeweils in zufälliger Abfolge dargeboten. Weder in den Vorversuchen noch im Hauptexperiment hat irgendeine der (anschliessend befragten) Versuchspersonen bemerkt, dass sie zweimal dieselben Aufgaben bearbeitet hatte. Zudem zeigten

¹⁸ Der Betrag von 10% des Gesamtwertes hat sich für One-Reason Decision Making als favorabel erwiesen (Läge, Hausmann & Christen, 2005).

sich in den Auswertungen auch keinerlei systematischen Reihenfolgeeffekte. Insofern kann die Reihenfolge der Bearbeitung als unkritisch angesehen werden.

Dieses Experimentaldesign liefert folgende Daten zu den einzelnen Fragenstellungen: Die *Hauptfragestellung* wird – neben einigen Häufigkeitsvergleichen im Hinblick auf die Diversifikation des Portfolios – durch die *Konkordanz in den Antwortmustern* jeder einzelnen Versuchsperson beantwortet: Herangezogen werden dazu diejenigen Optionsscheine, in die eine Person in einer bestimmten Aufgabe investiert hat (exakte Höhe der Investition und Betrag in die Sparbüchse werden dabei vernachlässigt). Liegen zur identischen Aufgabe in Bedingung 1 und 2 Investitionen in dieselben Optionsscheine vor, so gilt das Entscheidungsverhalten als gleich. Liegen Investitionen in unterschiedliche Optionsscheine vor (Art und/oder Anzahl), so gilt das Verhalten als unterschiedlich. Im paarweisen Vergleich aller 20 Aufgaben aus den zwei Bedingungen pro Person ergibt sich somit ein Konkordanzwert zwischen 0 und 20 gleichen Investitionen (auch ausdrückbar als Prozentwert bezogen auf das Maximum von 20 Aufgaben). Auf analoge Weise können sämtliche Versuchspersonen miteinander verglichen werden und zwar nicht nur mit je zwei unterschiedlichen Bedingungen, sondern auch innerhalb derselben Bedingung. Auf diese Weise entstehen bei 35 Versuchspersonen $35 \times 34 / 2$ Konkordanzwerte innerhalb aller Bedingungen 1, genau so viele innerhalb aller Bedingungen 2, 35×34 Konkordanzwerte zwischen unterschiedlichen Bedingungen von verschiedenen Personen und 35 Konkordanzwerte für die unterschiedlichen Bedingungen bei jeweils derselben Person.

Von einem systematischen Einfluss der Bedingungen auf das Entscheidungsverhalten soll dann gesprochen werden, wenn die 35 Konkordanzwerte für die unterschiedlichen Bedingungen bei jeweils derselben Person näher an den Konkordanzwerten innerhalb derselben Bedingung liegen als an einem identischen Investitionsmuster.

Fragestellung 2 (stärkeres Vertrauen auf das beste Argument vs. stärkeres Vertrauen auf Informationsintegration) kann anhand der Höhe der investierten Beträge beantwortet werden. Dazu berechnen wir die durchschnittliche Höhe desjenigen Betrags, der je Aufgabe in den Optionsschein mit dem validesten positiven Cue investiert wurde, und vergleichen ihn mit dem Betrag, der durchschnittlich auf den Optionsschein mit der besten Ratio gemäss Dawes' Rule gesetzt wurde. (Haben zwei Optionsscheine dieselbe Ratio, so gilt die Validität als Tie-Breaker.) Diese Auswertung bietet sich nur für Bedingung 1 an, da in Bedingung 2 sehr unterschiedliche Informationsmuster erwartet werden, die strukturell bereits durch Subgruppen hinsichtlich der Zielvariablen geprägt sein könnten.

Fragestellung 3 (Such- und Stoppregel) lässt sich über die Beobachtung options- oder attributweiser Informationssuche beantworten¹⁹. Tendiert eine Person zum *One-Reason Decision Making*, so wird sie mehrheitlich attributweise suchen (und zwar beginnend mit dem Cue, der der Person hinsichtlich Validität und Diskriminationsrate am günstigsten erscheint). Findet sie beim

¹⁹ Diese Form optionsweiser oder attributweiser Informationssuche wird bei Fasolo, McClelland & Lange (2005) im marktpsychologischen Kontext diskutiert.

ersten Cue einen Hinweis auf eine positive Entwicklung eines Optionsscheins, so wird sie die Informationssuche beenden und ihr Geld auf diesen Optionsschein setzen. (Findet sie positive Hinweise für mehrere Optionsscheine, so splittet sie entweder den Betrag oder setzt auf den mit der validesten positiven Information.) Nur wenn sie keinen positiven Hinweis findet, kauft sie einen weiteren Cue.

Ein Alternativen testendes Vorgehen im Sinne des *Satisficing* provoziert eine optionsweise Suchstrategie: Zunächst wird die Versuchsperson alle fünf Cue-Informationen zu einem (aufgrund von Validität und Diskriminationsrate geeignet erscheinenden) Optionsschein anschauen. Ist sie mit der Gesamtprognose auf den fünf Attributen zufrieden, so investiert sie in diesen Optionsschein. Ansonsten fährt sie mit der Informationssuche fort, indem sie zu einem zweiten Optionsschein wiederum sämtliche Attribute anschaut.

Wer hingegen eine optimale Integration sämtlicher Informationen vornehmen möchte (die „klassische“ Strategie zu Entscheiden), der sollte möglichst viel Information anschauen. Da es im Sinne der Vermögensmaximierung nicht zielführend ist, alle 25 Cue-Informationen zu kaufen, erwarten wir eine *Scan-und-Test-Strategie*: Zunächst wird ein Cue gekauft, um über alle Optionen zumindest je eine Information zu haben. Diejenigen Optionsscheine, zu denen eine negative Prognose erscheint, werden nicht weiter berücksichtigt (im Sinne von *elimination by aspects*). Zu denjenigen Optionsscheinen hingegen, zu denen sich eine positive Information findet, werden alle weiteren Informationen gekauft, um diese integrieren zu können²⁰. Diese Scan-und-Test-Strategie als Optimierung einer vollständigen Informationssuche ist in Abbildung 3.3 illustriert.

Für *Fragestellung 4* werden als Indikatoren der Risikobereitschaft einer Person folgende zwei Kennwerte aus Bedingung 1 herangezogen: Der Betrag, den eine Person durchschnittlich in die Sparbüchse steckt, sowie die Häufigkeit, mit der in einer Aufgabe in mehr als einen Optionsschein investiert wird. Niedrige Kennwerte sprechen für hohe Risikobereitschaft, hohe Kennwerte für Risikoaversion. Gleichzeitig gilt: Wer risikoaversiv ist, der will wissen, was er tut. Wir erwarten deshalb, dass risikoaversive Personen in Bedingung 2 ihr Informationssuchverhalten so ausrichten, dass sie bei einer Investition in einen Optionsschein hinsichtlich des Erfolgs möglichst sicher sind. Das bedeutet, sie suchen optionsweise. Aufgrund der Kategorialität der Suchvariable operationalisieren wir den erwarteten Zusammenhang zwischen Suchstrategie und Entscheidungsverhalten folgendermassen: Personen, die in Bedingung 2 mehrheitlich optionsweise suchen, zeigen in den beiden Kennwerten in Bedingung 1 systematisch höhere Werte als Personen, die vor allem Cue-weise nach Informationen suchen. Bestätigt sich dieser Zusammenhang, so ist das ein Hinweis darauf, dass die allgemeine Entscheidungsstrategie bereits Rückwirkungen auf die Informationssuche hat.

²⁰ Wahlweise können auch alle Informationen zu Optionsscheinen gekauft werden, bei denen der zuerst angeschaute Cue nicht anwendbar waren.

3.3 Resultate

3.3.1 Generelles Anlageverhalten: Bedingung mit vorgegebener Information

Betrachten wir zuerst die Bedingung mit vorgegebener Informationsbasis. Die Versuchspersonen konnten sämtliche offen auf dem Bildschirm vorliegenden Hinweise für ihre Entscheidungen berücksichtigen und verarbeiten (5x5-Datenmatrix). Die normative Strategie (im Sinne des *homo oeconomicus*) besteht jeweils darin, die gesamten zur Verfügung stehenden Geldmittel (10'000 Franken) auf diejenige Option mit der höchsten Eintretenswahrscheinlichkeit des Gewinns zu setzen. Anstelle der exakten Verrechnung der Validitätswerte konnte diese beste Option auch ungewichtet, d. h. mit einer reinen Auszählstrategie (Tallying im Sinne der Dawes' Rule), ohne großen Aufwand approximiert werden. Diese Strategie hätte beim zugrunde liegenden Datenmuster zu 18 von 20 Treffern geführt und so das Portfolio entsprechend maximiert.

Dessen ungeachtet gab es nur zwei von 35 Personen, die in sämtlichen 20 Aufgaben mit vorgegebener Information konsequent jeweils auf genau eine Option setzten. Insgesamt waren es im Schnitt lediglich 45% der 20 Aufgaben ($M = 9.09$, $SD = 6.32$) pro Versuchsperson, bei denen genau auf eine Option gesetzt wurde. Die Versuchspersonen zeigten vielmehr ein diversifiziertes Anlageverhalten ihres Budgets (Diversifikationsstrategie), indem sie ihre zur Verfügung stehenden Geldmittel auf mehrere Optionen verteilten. Bei 37% aller 20 Aufgaben wurden die Geldmittel auf zwei, bei 15% auf drei, bei 2% auf vier Optionen und nur in einem einzigen Fall (0.2%) auf alle fünf Optionen verteilt angelegt.²¹

3.3.2 Generelles Suchverhalten: Bedingung mit aktiver Informationssuche

Bei der Bedingung mit aktiver Informationssuche mussten die Versuchspersonen Informationen selber aktiv aufdecken. Indem sie entweder auf eine Option oder einen Indikator klickten, wurden spalten- bzw. zeilenweise jeweils fünf Hinweise auf einmal sichtbar. Anhand dieses Ausprägungsmusters konnten sie sich bereits zu diesem Zeitpunkt für eine oder mehrere Optionen entscheiden oder nach zusätzlichen Informationen suchen. Im Schnitt wurde 2.29 mal nach Informationseinheiten (à fünf Hinweise) gesucht ($SD = 0.64$). Damit waren im Mittel 46% des gesamten Informationsmusters analog der vorgegebenen Bedingung (inkl. der nicht diskriminierenden Informationen) für die eigentliche Entscheidung sichtbar. Lediglich in 2% der Aufgaben wurde geraten (keine Informationen gekauft) und in 3% der Aufgaben wurden (mit fünf Klicks) sämtliche Informationen aufgedeckt, was dem vollständigen Informationsmuster gemäss der vorgegebenen Bedingung entsprach.

Die Versuchspersonen konnten auch in der Bedingung mit aktiver Informationssuche ihr Budget (abzüglich der Informationskosten) auf eine oder mehrere der fünf Optionen setzen.²² Im Vergleich zur Bedingung mit vorgegebener Information setzten die Versuchspersonen mit durch-

²¹ Das Geld, welches in die Sparbüchse gelegt wurde, ist hier nicht berücksichtigt.

²² Grundsätzlich konnten die Versuchspersonen in der Such-Bedingung insgesamt weniger Profit machen, weil sich ja das Budget zum Anlegen in Optionen durch die Suchkosten verringerte.

schnittlich 57% der insgesamt 20 Aufgaben ($M = 11.37$, $SD = 6.77$) signifikant häufiger ihr gesamtes verbliebenes Geld auf nur eine Option ($t(34) = 2.53$, $p = .008$, einseitig). Wenn dieselben Personen aktiv nach Informationen suchen, diversifizieren sie also signifikant weniger. Und es gibt bereits 5 von 35 Personen, die in sämtlichen 20 Aufgaben mit vorgegebener Information konsequent auf genau eine Option setzten. Bei 27% aller 20 Aufgaben wurden die Geldmittel auf zwei, bei 11% auf drei und bei 4% auf vier Optionen verteilt angelegt.

3.3.3 Übereinstimmung im Anlageverhalten zwischen und innerhalb der Versuchspersonen (Hauptfragestellung)

Trifft eine Person nun dieselbe Entscheidung, wenn sie einmal aktiv und sequenziell nach Informationen sucht, im Vergleich zu einer bereits vollständig vorliegenden Informationslage? Dazu wird in der oben angegebenen Weise die Antwortkonkordanz herangezogen. In den 20 Aufgaben der Bedingung mit vorgegebener Information beträgt die durchschnittliche Übereinstimmung der paarweisen Vergleiche gemittelt über die 35 Versuchspersonen 0.33 ($SD = .15$), in den 20 Aufgaben der Bedingung mit aktiver Informationssuche mit 0.22 ($SD = .11$) signifikant tiefer ($t(38) = 2.44$, $p < .01$, einseitig). Informationssuche führt also zu größerer Varianz im Entscheidungsverhalten.

Als zentraler Vergleich interessiert hier aber die intrapersonale Konkordanz der Versuchspersonen zwischen den beiden Bedingungen. Als minimales Kriterium, um von einem systematischen Unterschied zwischen den zwei Bedingungen sprechen zu können, war gesetzt worden, dass die intraindividuelle Konkordanz näher am Mittelwert der interindividuellen Konkordanz (innerhalb derselben Bedingung) liegt als am Maximalwert von 1.00. Gemessen an der interindividuellen Konkordanz in Bedingung 1 darf die intraindividuelle Konkordanz also nicht größer als 0.66 ausfallen.

Die gemessene intraindividuelle Übereinstimmung beträgt im Durchschnitt 0.31 ($SD = .18$), wobei die individuellen Werte zwischen 0.85 und 0.00 schwanken. Die Versuchspersonen treffen somit in identischen Aufgabensets im Mittel in knapp 70% der Fälle *unterschiedliche* Wahlen, je nachdem ob ihnen die Informationen bereits auf dem Tablett präsentiert werden oder ob sie selber aktiv und sequenziell nach den verdeckten Informationen suchen müssen. Da der gemessene Wert sogar noch um zwei Prozentpunkte unter der interindividuellen Konkordanz aus Bedingung 1 liegt, kann man also sagen, dass der Unterschied, den die Informationssuche ausmacht, im Mittel gleich gross ist wie das Austauschen eines Individuums durch ein beliebiges anderes.

Aufgrund der grossen Streuungen in den Messwerten soll eine zusätzliche Statistik das Ergebnis beleuchten: Die Matrix der individuellen Konkordanzwerte wurde einer Nonmetrischen Multidimensionalen Skalierung (NMDS) unterzogen. Eine Projektion in einen niedrigdimensionalen Raum sollte im Falle eines geringen Einflusses der Bedingungen dazu führen, dass die Punkte für die beiden Bedingungen derselben Versuchsperson nah zueinander zu liegen kommen (weil sie relativ zu den anderen Konkordanzwerten dann sehr hoch ausfielen). Abbildung 3.4 zeigt eine solche Vorhersage anhand einer hypothetischen Karte für den zweidimensionalen Raum. Im Falle einer Bestimmung der Konkordanz durch die beiden Bedingungen sollten – im krassen

Gegensatz zu Abbildung 3.4 – alle Punkte der Bedingung 1 in einen Teil des Raumes kommen und alle Punkte der Bedingung 2 in einen anderen.

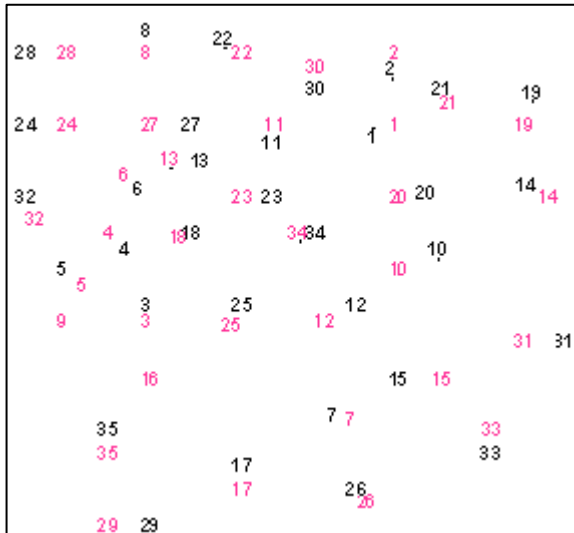


Abbildung 3.4: Hypothetische NMDS-Struktur, falls die Konkordanz innerhalb der Personen die zwischen den Bedingungen dominiert. Schwarz (fett): vorgegebene Information; magenta: aktive Informationssuche

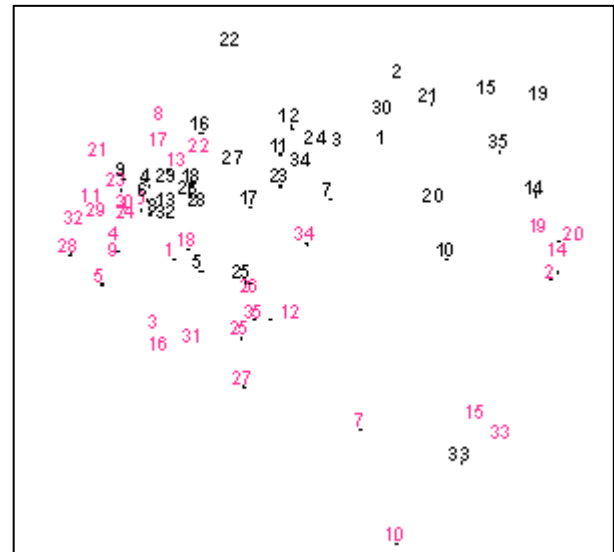


Abbildung 3.5: Resultierende NMDS-Lösung aus der Matrix der paarweisen Konkordanzwerte. Schwarz (fett): vorgegebene Information, magenta: aktive Informationssuche

Das Resultat (Abbildung 3.5) spricht deutlich für die zweite Hypothese: Die Nachbarschaftshypothese bestätigt sich nur für eine geringe Zahl von Versuchspersonen (z.B. die Personen 14, 33, 25 oder 5). Vielmehr ist der Raum durch eine deutlich erkennbare Trennung zwischen den beiden Bedingungen bestimmt (mit Ausnahme von Person 33 lassen sich die beiden Punktverteilungen zwar nicht als Cluster, doch als einfach zusammenhängende Objekte beschreiben). Nicht überraschend ist ferner, dass die Bedingung mit vorgegebener Information (im zentralen und rechten oberen Bereich) wegen der höheren Konkordanz von durchschnittlich 0.33 weniger Raum beansprucht als die weiter streuende Bedingung mit Informationssuche (mittlere Konkordanz 0.22).

3.3.4 Vertrauen in Information (Fragestellung 2)

Mit Fragestellung 2 soll geklärt werden, ob insgesamt stärkeres Vertrauen in das beste (= valide) Argument besteht (non-kompensatorisches One-Reason Decision Making) oder ob die Versuchspersonen mehr auf die Summe der positiven Argumente setzen (Integration der positiven Ausprägungen gemäss Dawes' Rule). Das vollständige Cue-Muster in Bedingung 1 lässt in 10 Aufgaben eine klare Unterscheidung zwischen den beiden Strategien zu, da hier die höchste Validität und die grösste Anzahl positiver Argumente auf verschiedene Optionen fielen. In diesen 10 Aufgaben setzten die Versuchspersonen mit durchschnittlich 5357 Franken ($SD = 3556$) signifikant mehr Geld auf die gemäss Dawes' Rule beste Option (MRDM) als auf die Option mit der validesten Cue-Ausprägung (ORDM) ($M = 946$ Franken, $SD = 1955$) ($t(349) = 18.23$, $p <$

.001). Das Verhalten der Versuchspersonen zeigt deutlich, dass ohne aktive Informationssuche nur sehr wenig Tendenz zum ORDM besteht.

Tabelle 3.1: Versuchspersonen klassifiziert nach der Art ihrer dominierenden Suchstrategie: optionen- bzw. spaltenweise Suche (Opt), attribut- bzw. zeilenweise Suche nach Cues (Cue) und Scan-und-Test-Strategie (zuerst attribut- und dann optionenweise Suche; ScT). Diejenigen Personen, welche die Cue-Strategie mehrheitlich nach dem ersten positiv diskriminierenden Cue stoppten, sind als One-Reason Decison Maker (ORDM) bezeichnet. Mit einem Stern (*) bezeichnete Vpn wiesen vereinzelt Durchgänge auf, in denen überhaupt keine Information gesucht wurde (reine Ratestrategie). Die mit zwei Sternen (**) bezeichneten Vpn wiesen vereinzelt oder gehäuft Durchgänge auf, in denen mindestens einmal attributweise gesucht wurde, nachdem bereits mindestens einmal eine Option angeklickt wurde.

Informationssuche mehrheitlich optionen-orientiert (Opt)				
Vp No.	Opt	Cue	ScT	Klassifikation
3	20	0	0	Opt
24	20	0	0	Opt
29	20	0	0	Opt
32	20	0	0	Opt
11	19	0	1	Opt
12	19	0	1	Opt
17*	19	0	0	Opt
21	18	2	0	Opt
25*	18	0	0	Opt
35**	18	0	0	Opt
2**	17	0	0	Opt
10**	16	0	0	Opt
14*	16	0	0	Opt
31**	16	2	0	Opt
19**	15	2	2	Opt
27*	15	0	0	Opt
Informationssuche mehrheitlich attribut-orientiert (Cue und/oder ScT)				
Vp No.	Opt	Cue	ScT	Klassifikation
1	0	20	0	Cue
4	0	20	0	Cue (ORDM)
7	0	20	0	Cue
28	0	20	0	Cue
30	0	20	0	Cue
18*	0	19	0	Cue (ORDM)
34	0	19	1	Cue
22	0	18	2	Cue
23	0	18	2	Cue
9**	0	12	7	Cue & ScT
26**	0	8	10	Cue & ScT
13	0	0	20	ScT
5**	0	2	17	ScT
15**	3	0	16	ScT
33**	2	1	16	ScT
Mischstrategien und andere Strategien				
Vp No.	Opt	Cue	ST	Klassifikation
6	14	5	1	Opt & Cue
16**	7	12	0	Opt & Cue
20**	8	0	1	Umkehrung der Scan-und-Test-Strategie
8**	4	0	4	keine klassifizierbare Strategie

3.3.5 Interindividuelle Unterschiede im Suchverhalten (Fragestellung 3)

Bei den Unterschieden im Suchverhalten ist zwischen Cue-orientierter (attributweiser) Informationssuche (bei ORDM mit Abbruch nach der ersten positiven Information), Optionsschein-orientierter (optionsweiser) Suche (Indikator für eine Satisficing-Strategie) und einer Scan-und-Test-Strategie (zuerst attributweise, anschliessend optionsweise) als MRDM zu unterscheiden. Bezüglich dieser Strategien bilden die Versuchspersonen ziemlich konsistente Verhaltensmuster aus (vgl. Tabelle 3.1): 9 Personen (26%) lassen sich einer attributweisen Suchstrategie zuordnen, 16 (46%) einer optionsweisen und vier Personen (11%) verfolgten eine Scan-und-Test-Strategie. Bei zwei Personen (6%) lässt sich eine bevorzugt cue-orientierte Informationssuche beobachten, mit einer Mischung aus attributweiser und einer Scan-und-Teststrategie. Zwei Personen (6%) zeigten eine Mischstrategie zwischen attribut- und optionsweiser Suche, Person 20 eine Umkehrung der Scan-und-Test-Strategie (attributweise Suche nachdem bereits mindestens eine Option aufgedeckt wurde) und Person 8 ein Verhalten, welches in keines der oben beschriebenen Muster passte.

Von den neun Versuchspersonen, die mehrheitlich eine konsequent Cue-orientierte Suche zeigten, stoppten allerdings nur zwei Personen mehrheitlich nach dem ersten Cue, der zu einem der Optionsscheine einen positiven Hinweis erhielt. One-Reason Decision Making ist also eine relativ seltene Strategie, während 7 der 9 Personen ihre Entscheidung durch mindestens ein zweites Argument absichern. Gemäss weiter oben eingeführter Operationalisierung ordnen wird diese sieben Probanden sowie die vier Personen der Scan-und-Test-Strategie, einer Informationsintegration (MRDM) zu. (Man kann auch noch die Vpn 9 und 26 hinzuzählen, je nachdem, wie strikt man die Operationalisierung sieht.) Satisficing (= optionsweise Suche) erwies sich mit 16 Personen als die beliebteste Vorgehensstrategie in diesem Setting. Keine der Versuchspersonen deckte übrigens alle Cues auf, um vollständige Informationsintegration nach Bedingung 1 vornehmen zu können.

3.3.6 Zusammenhang zwischen Anlageverhalten und Suchstrategien (Fragestellung 4)

Fragestellung 4 soll darüber Aufschluss geben, ob ein bestimmtes Anlageverhalten mit einer bestimmten Suchstrategie (attribut- vs. optionsweise Suche) zusammenhängt. Dazu wird die Risikobereitschaft (operationalisiert durch die Höhe des getätigten Betrags in die Sparbüchse und die Verteilung auf mehrere Optionen) als abhängige Variable zwischen den beiden Suchstrategietypen verglichen (siehe Klassifizierung bei Fragestellung 3). Die 15 Personen, die in Bedingung 2 bevorzugt attributweise aufdeckten, zeigten sich insofern risikosuchend, indem sie vom zu investierenden Betrag sowohl in Bedingung 2 ($M = 143$; $SD = 231$) als auch in Bedingung 1 ($M = 257$; $SD = 328$) vergleichsweise wenig in die Sparbüchse legten (vgl. Abbildung 3.6). Demgegenüber riskierten die 16 Personen mit bevorzugt optionsweiser Informationssuche wie erwartet weniger und steckten deutlich mehr Geld in die Sparbüchse: Durchschnittlich 438 Franken ($SD = 714$) in Bedingung 2 mit aktiver Informationssuche und mit 863 Franken ($SD = 1258$) in Bedingung 1 mit vorgegebener Informationssuche sogar signifikant mehr als die cue-orientierte Vergleichsgruppe ($t(29) = 1.81$, $p = .041$, einseitig). Gegenüber der Gruppe mit attri-

butorientierter Suche ist dieser Betrag also etwa dreimal so gross, und zwar auch in Bedingung 1, in der ja keinerlei Unterschiede in der jeweils vorliegenden Information bestanden.

Wer optionsweise nach Informationen sucht, der verteilt seinen Betrag pro Aufgabe in beiden Bedingungen zudem tendenziell auf mehr als einen Optionsschein (in Bedingung 1 auf durchschnittlich 1.85 Optionen ($SD = 0.54$) und in Bedingung 2 auf 1.80 ($SD = 0.77$)) (vgl. Abbildung 3.7). Cue-orientierte Informationssucher hingegen investieren grundsätzlich weniger häufig in mehrere Optionen, in Bedingung 1 mit vorgegebener Informationssuche zwar nur unscheinbar weniger ($M = 1.60$; $SD = 0.57$), in Bedingung 2 mit aktiver Informationssuche aber signifikant weniger ($M = 1.34$; $SD = 0.44$) als die cue-orientierte Vergleichsgruppe ($t(29) = 2.06$, $p = .026$, einseitig). Wiederum erweisen sich Personen mit einer an den Optionen orientierten Suchstrategie als die risikoaversiveren.

Zusätzlich zu beobachten ist übrigens, dass bei beiden Massen bei beiden Personengruppen die jeweils in Bedingung 2 (aktive Informationssuche) gezeigten Werte für eine grössere Risikofreudigkeit sprechen. Es scheint also so zu sein, dass die aktive Informationssuche die Konfidenz erhöht, obwohl im Endeffekt weniger Information zur Verfügung steht als beim vollständigen Information Board in Bedingung 1!

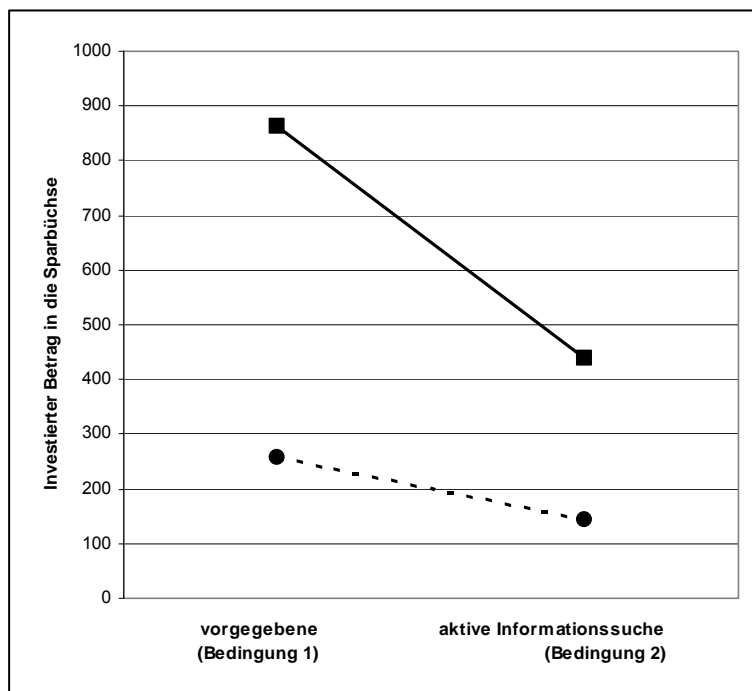


Abbildung 3.6: Risikobegrenzung: Geld in die Sparbüchse. Bedingung 1: vorgegebene Information, Bedingung 2: Information gesucht. Personen mit optionsweiser Suchstrategie (durchgezogene Linie) legen ein Mehrfaches dessen in die Sparbüchse, was Personen mit attributweiser Informationssuche (gestrichelte Linie) sparen.

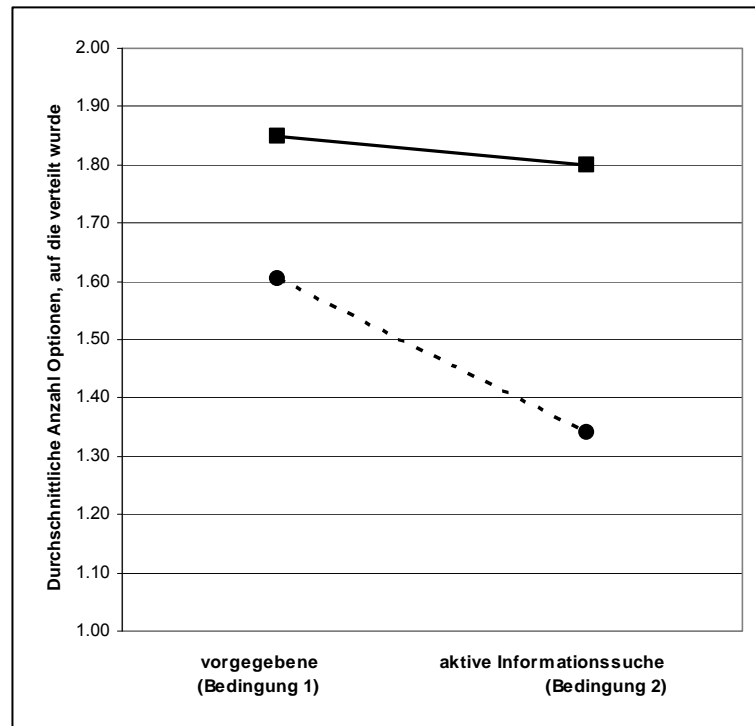


Abbildung 3.7: Risikobegrenzung: Streuung des Portfolio. Bedingung 2: vorgegebene Information, Bedingung 1: Information gesucht. Personen mit optionsweiser Suchstrategie (durchgezogene Linie) investieren ihr Geld in durchschnittlich mehr verschiedene Optionsscheine pro Durchgang als Personen mit attributweiser Informationssuche (gestrichelte Linie) dies tun.

3.4 Diskussion

Das vorliegende Experiment hat zu allen vier Fragestellungen aufschlussreiche Antworten erbracht. Am wichtigsten ist zweifellos der Unterschied, den der Faktor „Informationssuche“ ausmacht. Ob mit bereits (komplett) vorliegender Information entschieden (Bedingung 1) oder hingegen zuerst die Information aktiv und sequentiell aufgedeckt (Bedingung 2) werden muss, macht einen deutlichen Unterschied. Bei der ohnehin grossen interindividuellen Varianz im Anlageverhalten sagen zu müssen, dass die Informationssuche intrapersonal so viel ausmacht wie das Ersetzen dieser Person durch eine beliebig andere, ist ein recht deutliches Resultat. Gesetzt den Fall, dass Informationssuche im alltäglichen Entscheiden den Standardfall darstellt, wäre daraus die Konsequenz zu ziehen, dass dieser Faktor in der künftigen experimentellen Forschung nicht weiter außer Acht gelassen werden darf, da Entscheidungen mit oder ohne aktive Informationssuche offensichtlich zwei unterschiedliche Situationsklassen darstellen. Ein vorgegebenes vollständiges Information Board stellt dann eben keine gute Annäherung an die tatsächlichen „unabhängigen Variablen“ des menschlichen Verhaltens dar.

Die Resultate zu den Fragestellungen 2 bis 4 erscheinen uns im Vergleich dazu untergeordneter Natur zu sein. Ihre Bedeutung sehen wir eher im Sinne von Indizien, die jedoch weitestgehend an die ganz konkrete Vorlage dieses Experimentes gebunden sind. Deswegen sehen wir von ei-

ner zu starken Verallgemeinerung ab, zumal das 5x5-Design eine Erweiterung gegenüber den bisherigen Experimentalstrukturen in der Forschung zu Urteilsheuristiken darstellt. (Gerade die Forschung zum ORDM kennzeichnet sich ja durch Zweialternativen-Designs.) Die Befunde sollten jedoch dazu führen können, weitere experimentelle Studien anzustossen, welche die hier als Nebenfragestellungen behandelten Aspekte dann stärker in den Mittelpunkt rücken und das Experimentaldesign ganz darauf ausrichten. Ausserdem wird es nötig sein, die grundsätzliche Frage zwischen One-Reason Decision Making und Informationsintegration anhand unterschiedlichster Umwelten zu erforschen. Das hier präsentierte Design enthielt bewusst solche Aspekte, die für ORDM favorabel sind (extrem hohe Validitäten, optimales Suchkosten-Gewinnverhältnis; Mehr-Alternativen-Design²³) als auch solche, die die Anwendung der Dawes' Rule ermöglichen (z.B. hohe Diskriminationsrate, so dass ausreichend Argumente zum Abzählen zur Verfügung standen, oder eine klare Indikation mit farbigen Pfeilen auf dem Information Board). Die Möglichkeit zur optionsweisen Suche wiederum (wodurch sich dieses Experiment beispielsweise vom Standard der Information-Boards zur Take-the-Best-Heuristik unterscheidet) macht ein sequentielles Alternativen-testendes Verfahren im Sinne des Satisficing möglich. Dass gerade diese Variante von der Mehrzahl der Versuchspersonen gewählt wird, sollte der experimentell gegenüber der vollständigen Informationsintegration und der ORDM-Forschung doch weit im Hintergrund befindlichen Forschung zur sequentiellen Alternativentestung im Sinne Simons den Rücken stärken. Üblicherweise wird dieses Verfahren ja für notwendigerweise sequentiell vorzunehmende Informationssuche diskutiert (wie etwa Wohnungssuche oder Partnerwahl). Doch auch in einem Design wie diesem scheint die vollständige Betrachtung von Alternativen gegenüber dem Abscannen des Gesamttraums durch einen geeignet erscheinenden Cue eine valable Alternative darzustellen.

Daraus die Folgerung abzuleiten, hier werde notwendigerweise so etwas wie Satisficing vorgenommen, wäre jedoch im Hinblick auf die Anlage dieses Experimentes überzogen: Erstens wurde nicht exakt kontrolliert, inwieweit überhaupt nach der ersten hinlänglich gut erscheinenden Alternative der Suchprozess abgebrochen wurde, und zweitens konnten die Personen natürlich jederzeit auf einen zuvor betrachteten Optionsschein zurückkommen.

Ohne Informationssuche, so zeigt die Auswertung zu Fragestellung 2, besteht nur ein recht geringes Vertrauen in den Cue mit der höchsten Validität. Spontan scheint die Dawes' Rule doch eine grössere Anziehungskraft auf die Versuchspersonen auszuüben. Die Schlussfolgerungen aus diesem Befund (der sich mit dem von Läge & Daub, 2006, deckt) sollten jedoch unbedingt auf die Situationsklassen „Entscheiden unter vorgegebener Information“ begrenzt bleiben: Wenn man Menschen viel Information auf den Tisch legt, dann tendieren sie auch dazu, diese nutzen zu wollen. Die Strategiewertung zu Fragestellung 2 zeigt allerdings, dass bei der Mehrzahl der Versuchspersonen auch in Bedingung 2 wenig Neigung zum One-Reason Decision Making besteht. Obwohl diese Strategie aufgrund der sehr hohen Validitäten und der vernünftig hohen

²³ Es finden sich in der Literatur Hinweise darauf, dass Entscheidungssituationen mit mehr als zwei Optionen non-kompensatorisches Entscheiden fördern: Billings & Marcus (1983); Klayman (1985); Onken, Hastie & Revelle (1985); Sundström (1987).

Diskriminationsraten extrem erfolgversprechend ist (man braucht nur 1'000 Franken, gelegentlich einmal 2'000 Franken zu investieren, um einen Optionsschein mit mehr als 90% Erfolgswahrscheinlichkeit zu finden), wenden die allermeisten Personen (zumindest in dieser Umwelt) kein ORDM an.

Bemerkenswert erscheint uns im Hinblick auf Fragestellung 3 schließlich auch zu sein, dass die Personen überhaupt intraindividuell so strategisch vorgehen. Experimente zur Take-the-Best-Heuristik zeichnen sich beispielsweise mehrheitlich durch ein intraindividuell sehr heterogenes Vorgehen aus (vergleiche den Überblick bei Bröder, 2005, aber auch die dort noch nicht besprochenen Resultate von Läge, Hausmann & Christen, 2005; Läge, Hausmann, Christen & Daub, 2005 sowie von Läge & Daub, 2006).

Eine kurze Diskussion sind auch die Indizien zum Zusammenhang von Suchstrategie und Risikoeinstellung wert. Wer attributsweise den Alternativenraum abscannt, der entscheidet riskobewusster als derjenige, der optionsweise sucht. Wir dürfen Risikofreudigkeit als Persönlichkeitsmerkmal ansehen, also als eine Größe, deren interindividuelle Varianz nicht durch das Experiment induziert, sondern nur abgerufen wird. Deswegen rechtfertigt sich die Interpretationsrichtung, dass unterschiedliche Risikoeinstellungen zu unterschiedlichen Suchstrategien führen. Allerdings soll auch dieser Befund lediglich im Sinne eines Indizes gewertet werden: Das Experiment war nicht spezifisch darauf ausgelegt, Risikoeinstellung als unabhängige Variable zu operationalisieren. Gefunden haben wir eine Konkordanz in unterschiedlichen Verhaltensvariablen – der Rest ist Interpretation. Wer möchte, der kann aber natürlich gerne so weit gehen, die gefundenen Suchstrategien auf Finanzmärkte anzuwenden: Analyse eines einzigen Unternehmens beispielsweise, dafür aber in all seinen möglichen Facetten, versus einer Reaktion ganzer Aktienmärkte auf Einzel-Cues wie Leitzinsänderungen, Umweltkatastrophen oder Arbeitsmarktdaten mit den dabei involvierten Anlageentscheidungen ...

In realen Entscheidungsprozessen wie z.B. im Finanzmarkt haben Personen selten einen vollständigen Einblick in sämtliche relevante Informationen. Informationen sind nicht einfach präsent, sondern müssen meist erst aktiv und sequenziell gesucht werden. Informationssuche wird so zu einem zentralen Faktor im Entscheidungsprozess. Ressourcen wie verfügbare Zeit, Suchkosten und kognitive Verarbeitungskapazität beschränken die Informationssuche und damit die Möglichkeiten, überhaupt umfangreiche Information zur Verarbeitung vorliegen zu haben. Insofern war es mehr als an der Zeit zu überprüfen, ob der der Entscheidung vorangehende – von der klassischen Entscheidungsforschung bislang vernachlässigte – Informationssuchprozess als Marginalität abgetan werden darf. Die Ergebnisse dieses Experimentes sprechen eine eindeutige Sprache. Unser Fazit möchten wir deswegen mit einem Rückbezug auf die Hauptfragestellung ziehen: Mit der vorliegenden Untersuchung konnte gezeigt werden, dass der Faktor der Informationssuche das Entscheidungsverhalten wesentlich beeinflusst. Dieselben Personen treffen in identischen Entscheidungsaufgaben unterschiedliche Entscheidungen: Der Unterschied besteht lediglich darin, ob sie aktiv und sequenziell nach Informationen suchen müssen, oder ob sämtliche Informationen bereits offen auf dem Tisch liegen. Die Informationssuche ist ein zentraler

Faktor im Entscheidungsprozess. Somit ist keine Theorie in der Entscheidungsforschung vollständig, wenn sie nicht die Phase der Informationssuche berücksichtigt.

3.5 Literatur

- Billings, R. S. & Marcus, S. A. (1983). Measures of compensatory and noncompensatory models of decision behavior: Process tracing versus policy capturing. *Organizational Behavior and Human Performance*, 31, 331-352.
- Borges, B., Goldstein, D. G., Ortman, A., & Gigerenzer, G. (1999). Can ignorance beat the stock market? In G. Gigerenzer, P. M. Todd & the ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 59-72). New York: Oxford University Press.
- Bröder, A. (2005). *Entscheiden mit der „adaptiven Werkzeugkiste“: Ein empirisches Forschungsprogramm*. Lengerich, Pabst.
- Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, 34, 571-582.
- Edwards, W., & Slovic, P. (1965). Seeking information to reduce the risk of decision. *American Journal of Psychology*, 78, 188-197.
- Eisenführ, F., & Weber, M. (2003). *Rationales Entscheiden (Rational decision making)*. Berlin: Springer, 4th ed.
- Fasolo, B., McClelland, G. H., & Lange, K. A. (2005). The effect of site design and interattribute correlations on interactive web-based decisions. In C. P. Haugtvedt, K. Machleit, & R. Yalch (Eds.), *Online consumer psychology: understanding and influencing behavior in the virtual world* (pp. 325-344). London: Lawrence Erlbaum Associates.
- Fischhoff, B. (1975). Hindsight = foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, 1, 288-299.
- Frey, D. (1981). *Informationssuche und Informationsbewertung bei Entscheidungen*. Bern: Hans Huber.
- Fried, L. S., & Peterson, C. R. (1969). Information seeking: optional versus fixed stopping. *Journal of Experimental Psychology*, 80, 525-529.
- Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, 103, 650-669.
- Gigerenzer, G., & Goldstein, D. G. (1999). Betting on one good reason: The Take The Best heuristic. In G. Gigerenzer, P. M. Todd & the ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 75-95). New York: Oxford University Press.
- Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review*, 98, 506-528.
- Gigerenzer, G., Todd, P. M., & the ABC Research Group (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.
- Goldberg, J., & Nitzsch von, R. (2000). *Behavioral Finance: Gewinnen mit Kompetenz*. München: FinanzBuch Verlag, 2nd ed.
- Huang, L.-N. (2000). Examining candidate information search processes: The impact of processing goals and sophistication. *Journal of Communication*, 50, 93-114.
- Huber, O., & Macho, S. (2001). Probabilistic set-up and the search for probability information in quasi-naturalistic decision tasks. *Risk Decision and Policy*, 6, 1-16.

- Huber, O., Beutter, C., Montoya, J., & Huber, O. W. (2001). Risk-defusing behaviour: Towards an understanding of risky decision making. *European Journal of Cognitive Psychology*, 13, 409–426.
- Huber, O., Wider, R., & Huber, O. W. (1997). Active information search and complete information presentation in naturalistic risky decision tasks. *Acta Psychologica*, 95, 15–29.
- Johnson, M. D. (1989). The differential processing of product category and noncomperable choice alternatives. *Journal of Consumer Research*, 16, 300–309.
- Jungermann, H., Pfister, H.-R., & Fischer, K. (2005). *Die Psychologie der Entscheidung: Eine Einführung*. Heidelberg: Spektrum Akademischer Verlag (2. Auflage).
- Kahneman, D., Slovic, P., & Tversky, A. (1982) (Eds.). *Judgment under uncertainty: heuristics and biases*. Cambridge, MA: Cambridge University Press.
- Kerstholt, J. H. (1992). Information search and choice accuracy as a function of task complexity and task structure. *Acta Psychologica*, 80, 185–197.
- Kiehling, H. (2001). *Börsenpsychologie und Behavioral Finance: Wahrnehmung und Verhalten am Aktienmarkt*. München: Vahlen.
- Kirchler, E. M. (1999). *Wirtschaftspsychologie: Grundlagen und Anwendungsfelder der Ökonomischen Psychologie*. Göttingen: Hogrefe, 2nd ed.
- Kirschenbaum, S. S. (1992). Influence of experience on information-gathering strategies. *Journal of Applied Psychology*, 77, 343–352.
- Klayman, J. (1985). Children's decision strategies and their adaptation to task characteristics. *Organizational Behavior and Human Decision Processes*, 35, 179–201.
- Läge, D., & Daub, S. (2006). *Entscheiden mit mehreren Cue-Informationen: die Entscheidungsregel am Beispiel des Städteparadigmas von Gigerenzer & Goldstein (1996)*. AKZ-Forschungsbericht 35. Zürich: Angewandte Kognitionspsychologie.
- Läge, D., Hausmann, D., & Christen, S. (2005). *Wie viel bezahlen für eine valide Information? Suchkosten als limitierender Faktor der Informationssuche*. AKZ-Forschungsbericht 07. Zürich: Angewandte Kognitionspsychologie.
- Läge, D., Hausmann, D., Christen, S., & Daub, S. (2005). *Was macht einen „guten Cue" aus? Strategien der Informationssuche beim heuristischen Entscheiden unter Unsicherheit*. AKZ-Forschungsbericht 05. Zürich: Angewandte Kognitionspsychologie.
- Lee, M. D., & Cummins, T. D. R. (2004). Evidence accumulation in decision making: Unifying the “take the best” and the “rational” models. *Psychonomic Bulletin and Review*, 11, 343–352.
- Martignon, L., & Hoffrage, U. (1999). Why does one-reason decision making work? A case study in ecological rationality. In G. Gigerenzer, P. M. Todd & the ABC Research Group (Eds.), *Simple heuristics that make us smart*. (pp. 119–140). New York: Oxford University Press.
- McDougal, Y. B. (1995). Decision making under risk: risk preference, monetary goals and information search. *Personality and Individual Differences*, 18, 771–782.
- Mill, J. S. (1848). *Principles of Political Economy with some of their Applications to Social Philosophy*. London: Longmans.
- Onken, J., Hastie, R. & Revelle, W. (1985). Individual differences in the use of simplification strategies in a complex decision making task. *Journal of Experimental Psychology: Human Perception and Performance*, 11, 14–27.
- Roetheli, C. (2002). *Wann wissen wir genug? Stopp-Mechanismen bei komplexen Entscheidungen*. Berlin: Schalungsdienst Lange.

- Shefrin, H. (2000). *Beyond greed and fear: Understanding Behavioral Finance and the Psychology of Investing*. Boston, MA: Harvard Business School Press.
- Sheluga, D. A., Jaccard, J., & Jacoby, J. (1979). Preference, search, and choice: An integrative approach. *Journal of Consumer Research*, 6, 166–176.
- Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, 69, 99–118.
- Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, 63, 129–138.
- Simon, H. A. (1957). *Models of man: Social and rational*. New York: Wiley.
- Sundström, G. A. (1987). Information search and decision making: The effects of information displays. *Acta Psychologica*, 65, 165–179.
- Tversky, A., & Kahneman, D. (1974). *Judgment under uncertainty: Heuristics and biases*. *Science*, 185, 1124–1131.
- Unser, M. (1999). *Behavioral Finance am Aktienmarkt*. Bad Soden: Uhlenbruch.

Teil II - Quizshow-Szenario

Kapitel 4

Die Quizshow: Ein experimentelles Design zur Messung von individuellen Informationssuch-Strategien in einfachen Urteilsheuristiken

Der vorliegende Forschungsbericht beschreibt ein Experimentaldesign zur Informationssuche beim Entscheiden. Die Beschreibung beispielhaft damit durchgeführter Experimente findet sich bei Christen, Hausmann & Läge (2006).

Der implementierte Ansatz ermöglicht es, die Suchregel als Baustein einer Urteilsheuristik auf individueller Ebene genauer zu erforschen. Grundlage dazu ist eine freie Variierbarkeit von Validität und Diskriminationsrate probabilistischer Hinweiscues von Durchgang zu Durchgang. Dies ist realisiert, indem die Versuchspersonen als Kandidaten einer Quizshow schwierige Almanachfragen zu beantworten haben, zu deren Lösung sie (in Anlehnung an die Quizshow „Wer wird Millionär?“) Telefonjoker heranziehen können. Diese Telefonjoker zeichnen sich durch ihren unterschiedlichen Expertisegrad (Validität), aber auch ihre unterschiedliche telefonische Erreichbarkeit (Diskriminationsrate) aus. Die sequentielle Wahl der Joker gibt dann Aufschluss darüber, welcher Suchregel eine Person folgt.

Gleichzeitig wird über die Gestaltung des Quiz' (z.B. mit Erläuterungen der korrekten Antworten und mit einem motivierenden Score-System) die Aufmerksamkeit der Versuchspersonen von der Informationssuche selbst abgezogen. Dadurch, dass erstens ihr Fokus realitätsnah bei der Bearbeitung von immer wieder neuen Aufgaben liegt und zweitens die Telefonjoker immer wieder variieren, vermeidet das Design die sonst übliche Standardisierung hin zu repetitiven Experimentaldurchgängen und öffnet somit die Türe zu einer grösseren externen Validität.

4.1 Einleitung: Methodologische Überlegungen zur Messung des Entscheidungsverhaltens

Die klassische Entscheidungsforschung kennt zwei unterschiedliche methodische Zugangsweisen zur Überprüfung von kognitiven Prozessmodellen bei multiplen Attributinformationen: sogenannte *process tracing methods* und das *structural modeling* (vgl. Übersicht in Bröder, 2000a; Harte & Koele, 2001). Das Ziel der empirischen Modellüberprüfung in der Entscheidungsforschung besteht grundsätzlich darin, mit einem geeigneten diagnostischen Verfahren auf die verwendete Urteilsstrategie zurückzuschliessen bzw. ein bestimmtes Modell zu bestätigen und andere Modelle auszuschliessen.

Process Tracing (PT) versucht, die eigentlichen kognitiven Prozesse abzubilden (Payne, 1976). Neben der direkten Verhaltensbeobachtung wie dem *Lauten Denken* und der *Analyse von Blickbewegungen* wurde die Verwendung eines *Information Display Boards* immer wichtiger (Sundström, 1987). Eine solche Informationstafel besteht aus einer (meist am Computerbildschirm dargebotenen) Matrix mit (häufig in Spalten dargebotenen) Objekten und (entsprechend in den Zeilen dargebotenen) Attributen bzw. Cues. Die Zellen beinhalten die vorerst verdeckten Attributs- bzw. Cue-Ausprägungen, die erst durch einen entsprechenden Mausklick in meist freier Reihenfolge – aber immer sequenziell und aktiv – aufgedeckt werden müssen. In Anlehnung an die empirischen Untersuchungen von Payne, Bettman & Johnson (1993) wird dieses Verfahren auch *Mouselab* genannt. Mit der Informationstafel wurden ursprünglich die Suchtiefe, die Variabilität und das Muster der Informationssuche gemessen (Payne, Bettman & Johnson, 1988). Für eine kritische Betrachtung dieser Masse siehe Böckenholt & Hynan, 1994a; 1994b; Harte, 1995; Bröder, 2000a. Darüber hinaus können zusätzliche Variablen überprüft werden, wie z.B. der Zeitpunkt und die Dauer der Betrachtung einzelner Informationen, die Suchreihenfolge und Cue-Präferenz, die Präferenz von optionenweiser vs. attributsweiser Verarbeitung, der Zeitpunkt des Abbruchs der Informationssuche und entsprechend generell der Umfang der gesuchten Informationen.

Die Methode des *Structural Modeling (SM)* setzt am Endprodukt des Entscheidungsprozesses an, indem lediglich der Output (Präferenz, Urteil, Wahl) mathematisch modelliert wird. Gerechnet wird typischerweise eine multiple lineare Regression, in die die Attributwerte als Prädiktoren und der Output als Kriterium eingehen (vgl. Brehmer, 1994; Brehmer & Brehmer, 1988; Stewart, 1988). Der so erhaltene multiple Korrelationskoeffizient R dient als Mass der Modellanpassung und der Konsistenz des Entscheiders (vgl. Kritik von Bröder, 2005). Die Regressionsgewichte der einzelnen Attribute dienen als Mass der Cue-Verwendung und lassen auf eine non-kompensatorische oder eine kompensatorische Informationsverarbeitung schliessen. Eine grundsätzliche Problematik von SM ist die hohe Überlappung der Vorhersagen für unterschiedliche Modelle. Beim Städteparadigma von Gigerenzer & Goldstein (1996) treffen beispielsweise 92% der paarweisen Objektvergleiche dieselbe Modellvorhersage für TTB und für ein kompensatorisches Modell. Auf dieses Separierungsproblem hat Bröder (2000a, 2000b) aufmerksam gemacht. Für eine weitere kritische Betrachtung des SM siehe Bröder, 2005; Fiedler, 1978). Eine grund-

sätzlich auf dem SM-Ansatz beruhende Weiterentwicklung hat Bröder (2000a; 2005; Bröder & Schiffer, 2003b) mit einer regressionsbasierten Methode und einer Bayesianischen Modellselektion geliefert. Beide Methoden verlangen zuerst die Formulierung eines Fehlermodells, da reale Versuchspersonen selten eine 100% konsistente Strategie verfolgen (Bröder, 2005, S. 91 ff.).

PT und SM werden häufig als Gegensätze betrachtet. Dabei können sich PT und SM vielmehr ergänzen (Einhorn, Kleinmuntz & Kleinmuntz, 1979; Harte & Koele, 2001). SM kann beispielsweise als hypothesengenerierendes Verfahren eingesetzt werden, welches klare Vorhersagen zwischen zwei Modellen aufstellt, die dann mittels PT auf individueller Verhaltensebene überprüft werden können. Dabei werden zwei Objekte sowie deren Cue-Ausprägungen so modelliert, dass z.B. die Wahl des Objektes A nur aufgrund des einen Modells (z.B. TTB) und die Wahl des Objektes B mit anderen Modellen (z.B. durch eine kompensatorische Entscheidungsregel) erklärt werden kann (vgl. auch die Diskussion zum Separierungsproblem Bröder, 2000a).

Die Idee der zweistufigen empirischen Überprüfung konkreter Urteilsheuristiken ist der methodische Leitfaden für das im Folgenden berichtete Experimentaldesign: Im ersten Schritt wird eine Urteilsaufgabe kreiert, deren Cue-Gütemasse oder Cue-Ausprägungen eine zielgerichtete Hypothese erlauben, welcher Cue bzw. welches Objekt gewählt werden sollte, wenn denn die Suchstrategie bzw. die Urteilsheuristik korrekt angewendet würde (SM). Im zweiten Schritt wird das Such- und Entscheidungsverhalten der Versuchspersonen in dieser Aufgabe anhand des aufgedeckten Cue-Musters festgehalten und hinsichtlich der Hypothese überprüft (PT). Die Konsequenzen dieses Vorgehens sind folgende:

1. Die einzelnen Prozessbausteine von Heuristiken zum One Reason Decision Making (Such-, Stopp- und Entscheidungsregel) müssen innerhalb einer Aufgabe separat als Hypothese formuliert und überprüft werden. Erst wenn sämtliche drei Bausteine korrekt angewendet werden, kann von einer korrekten Anwendung der ORDM-Heuristik (z.B. TTB) gesprochen werden. Diese Sichtweise ist neu, entspricht aber der Forderung nach einer strengen empirischen Überprüfung der Modellvorhersagen (vgl. Bröder, 2000a; 2004).
2. Die Überprüfung einer Urteilsheuristik ist auf individueller Versuchspersonenebene durchzuführen (vgl. die Forderung von Bröder, 2005, S. 83), weil nur für den Einzelfall Verletzungen der Bausteine festgestellt und interpretiert werden können. Dies ist durchaus als Vorteil zu sehen, da nicht a priori ausgeschlossen werden kann, dass die Mehrzahl bzw. sämtliche Versuchspersonen keine einheitliche Strategie verfolgen.
3. Die Überprüfung muss anhand mehrerer Aufgaben wiederholt werden. Referenz wird die Anzahl gelöster Aufgaben pro Versuchsperson und nicht die Stichprobengröße. Dafür erübrigen sich aufwändige statistische Prüfverfahren. Häufig reicht ein Binomialtest für die Abgrenzung gegen die zufällige Verwendung von einzelnen Bausteinen oder die Klassifikation zu einem bestimmten Modell.
4. Daraus lässt sich statistisch nicht nur die Bevorzugung einer bestimmten Strategie nachweisen, sondern auch zeigen, wann eine Versuchsperson keine Strategie hat und nur im Hinblick auf

die überprüften Theorien zufälliges Verhalten zeigt (vgl. Kritik von Bröder, 2004, am herkömmlichen SM-Verfahren).

Für die adäquate empirische Überprüfung von Urteilsheuristiken müssen einige weitere Voraussetzungen erfüllt sein: So muss z.B. sichergestellt sein, dass die Versuchsperson für ihre Urteile ausschliesslich die im Experiment verwendete Informationsbasis verwendet, da sonst das Entscheidungsverhalten nicht valide gemessen werden kann. Bröder (2000a; 2005, S. 84) bezeichnet dies als Identifizierungsproblem. Es gibt zwei methodische Zugänge, um das Vorwissen auszuschalten: 1. Es werden fiktive Szenarien verwendet, hinter denen aber ein realer Datensatz stehen kann. 2. Es werden reale Szenarien verwendet und der Grad des Vorwissens wird systematisch miterhoben.

Um im zweiten Fall das Vorwissen einer Versuchsperson in einer konkreten Urteilsaufgabe zu messen, muss sie vor der eigentlichen Informationssuche eine Präferenz für eines der Objekte abgeben. Dieses wird mittels eines Konfidenzurteils abgefragt. Die Abfrage eines Konfidenzurteils war von Gigerenzer et. al. (1991) ursprünglich für das abschliessende Urteil vorgesehen, um die Strategien des Wissens und zufälligen Ratens vom eigentlichen induktiven Erschliessen trennen zu können. Probabilistische Inferenzen können nur dann herangezogen werden, wenn die Zielgrösse der Objekte unbekannt, die Cue-Validitäten und Cue-Ausprägungen hingegen bekannt sind. Das vermeintlich korrekte Urteil wird mit einer bestimmten Urteilssicherheit induktiv erschlossen.

Eine weitere Voraussetzung der empirischen Überprüfung ist das adäquate Verständnis des Cue-Konzeptes und seiner Gütemasse Validität (V) und Diskriminationsrate (D). Im weiter unten zu beschreibenden Experimentaldesign wird den Versuchspersonen die Validität mit den Begriffen „Zuverlässigkeit“ oder „Trefferrate“ erklärt und die Diskriminationsrate als „Erreichbarkeit“ oder „Anwendbarkeit“. Zudem ist das Cue-Konzept dahingehend modifiziert, als hinter den Cues nicht inferenzielle Merkmale einer Zielvariable stehen, sondern personifizierte Informanten (Cues). Dies lässt einerseits die Cues unabhängig(er) voneinander werden und andererseits erlaubt es die Ausdehnung des Ausprägungsbereichs für die Gütemasse V und D im experimentellen Design. Die Zuverlässigkeit (Expertisegrad, V) eines Informanten (Cue) kann damit frei zwischen der Zufallswahrscheinlichkeit und 1.00 schwanken, je nachdem wie viele korrekte Vorhersagen er in der Vergangenheit getroffen hat bzw. treffen wird.

Durch die Ausdehnung der Objektzahl kann die Überprüfung von TTB auf den Mehralternativenfall übertragen werden, was wiederum einen Einfluss auf die Basisrate hat. Die binäre Cue-Ausprägung kann beibehalten werden, indem die Merkmalsausprägung des personifizierten Cues (mit der entsprechenden Zuverlässigkeit) auf eines der Objekte zeigt (Beispiel: Cue *a* weist mit Zuverlässigkeit *p* auf Objekt *x*). Auch das Antwortverhalten des Cues bleibt binär, wenn der Versuchsperson lediglich rückgemeldet wird, ob ihr Urteil (und damit die Vorhersage des Cues) korrekt oder falsch war. Die Erreichbarkeit lässt sich dank diesen Massnahmen (personifizierte Cues) praktisch zwischen 0.00 und 1.00 variieren.

4.2 Anforderungen an das Versuchsdesign

Die wichtigste Anforderung an das neue Versuchsdesign besteht darin, eindeutig herauszufinden zu können, ob sich Personen zur Beurteilung der Güte eines Cues bei der Informationssuche eher auf die Validität (V) oder die Diskriminationsrate (D) oder auf eine Kombination dieser beiden Werte abstützen. Ausserdem soll damit die Möglichkeit geschaffen werden, den Grad der Konsistenz einzelner Versuchspersonen im Hinblick auf ihre bevorzugte Informationssuch-Strategie zu messen.

Eine zweite wichtige Anforderung an das Experimentaldesign ist, über ein möglichst realitätsnahes (Spiel-)Szenario zu verfügen, das die relevanten entscheidungspsychologischen Konzepte 'Validität', 'Diskriminationsrate' und 'Cue' in eine möglichst natürliche, d.h. alltagsnahe Darstellungsform bringt. Diese Konzepte sollen von den Versuchspersonen auf einfache Weise als natürliche Bestandteile der Versuchsumgebung erkannt werden und damit ohne langwierige, theoretische und komplizierte Erklärungen intuitiv verständlich sein.

4.2.1 Bestimmung der angewendeten Informationssuch-Strategie

Mit dem Experimentalszenario wollen wir fünf theoretisch postulierte Informationssuch-Strategien auf ihre Anwendungshäufigkeit hin überprüfen: Suche in der Reihenfolge absteigender Cue-Validität bzw. Diskriminationsrate, Success, Usefulness sowie „V + D“ (als zwar untaugliche, aber rechnerisch einfache Kombination aus Validität und Diskriminationsrate). Abhängig von der Gewichtung (bzw. Verrechnung) von V und D resultieren unterschiedliche, optimale Cue-Reihenfolgen und damit unterschiedliche, optimale Suchstrategien. Über die Auszählung der Anwendungshäufigkeiten der einzelnen Suchstrategien pro Versuchsperson können wir einerseits auf die individuelle Präferenz der untersuchten Suchstrategien und damit auf die individuelle Gewichtung von V und D schliessen und andererseits den Grad der Konsistenz messen, mit der die einzelnen Strategien angewendet werden.

4.2.2 Aktive Informationssuche in 'realer Welt'

Die Reihenfolge der einzelnen Informationssuch-Schritte ist für die Festlegung der individuellen Suchstrategie entscheidend und muss daher für die Experimentatoren post-experimental sichtbar und nachvollziehbar sein. Die Informationssuche darf deshalb nicht gedächtnisbasiert sein, sondern muss in der 'realen Welt' des Experimental-Settings erfolgen.

Besässe eine Versuchsperson ausreichend Vorwissen, auf das sie sich bei der Lösung der Aufgaben abstützen könnte, bräuchte sie nicht nach zusätzlichen Informationen zu suchen. Deshalb müssen die Aufgaben so konstruiert sein, dass in den allermeisten Durchgängen kein Vorwissen eingesetzt werden kann. Eine häufig gewählte Möglichkeit um Vorwissen auszuschliessen, bestünde darin, die Experimental-Umgebung realitätsfern auszugestalten. Im Fall des vorliegenden Experimental-Settings schlagen wir aber bewusst einen anderen Weg ein: Die Versuchspersonen sollen in eine Entscheidungssituation gebracht werden, in der es um ganz reale Aufgaben geht. Einen kleinen Teil der Aufgaben werden sie dabei mit ihrem Wissen erschliessen können, in den allermeisten Fällen jedoch sollen sie auf zusätzliche Informationsquellen angewiesen sein, die

ihnen in Form probabilistischer Cues angeboten werden. Die Cues sollen, wie in der Entscheidungsforschung üblich, in ihrer formalen Qualität durch Validität und Diskriminationsrate ausgezeichnet sein. Die Versuchspersonen haben aufgrund ihres natürlichen Vorwissens zu der jeweiligen Aufgabe und aufgrund der Cuequalitäten zu entscheiden, ob sie einen gewissen Aufwand (Kauf) betreiben wollen, um an eine oder mehrere der probabilistischen Informationen heranzukommen.

4.2.3 Standardisierung der realen Welt

Damit alle Versuchspersonen die gleichen Ausgangsbedingungen haben und damit eine Messung der individuellen Suchstrategien möglich ist, muss eine gewisse Standardisierung im Ablauf vorliegen. Diese Standardisierung erfolgt dadurch, dass die reale Welt (bzw. ein winziger Ausschnitt davon) in das Experiment eingebaut wird, allerdings so, dass die abrufbare Information fest vorgegeben und dadurch begrenzt ist. Damit ist nicht nur für standardisierte Versuchsbedingungen gesorgt, sondern gleichzeitig auch für die Möglichkeit, den Prozess der Informationssuche (retrospektiv) mitverfolgen zu können, indem die Reihenfolge des Informations-Abrufs festgehalten wird.

4.3 Umsetzung der Anforderungen

4.3.1 Schwierige Wissensfragen

Unter Berücksichtigung der bisher erwähnten Möglichkeiten und Anforderungen (Untersuchungsziel, Suchstrategien, realitätsnahes Setting, natürliche Präsentationsform der entscheidungspsychologischen Konzepte, Informationssuche in der realen Welt unter standardisierten Versuchsbedingungen, objektiv richtige Entscheidung, Ausschluss von Vorwissen) und inspiriert vom Konzept der bekannten TV-Quizshow "Wer wird Millionär?" entstand folgendes Experimental-Setting: Wir stellen den Versuchspersonen speziell ausgewählte, schwierige Wissensfragen, so dass sie die richtige Antwort nur in seltenen Fällen bereits wissen. Die Beantwortung der Fragen wird in den meisten Fällen nur mit Hilfe zusätzlicher Informationen möglich sein. Diese Informationen müssen aus probabilistischen Quellen (Cues) erschlossen werden. Für jede richtig beantwortete Frage erhalten sie einen kleinen Spielgewinn, der (im Gegensatz zu "Wer wird Millionär?") über das ganze Spiel konstant bleibt.

4.3.2 Telefonjoker

Als Hilfestellung bei der Beantwortung der Wissensfragen bietet das Setting pro Aufgabe sechs Telefonjoker (die Cues) an. Diese sind mit einer bestimmten Wahrscheinlichkeit (Diskriminationsrate) telefonisch erreichbar und tippen, falls sie erreicht werden, mit einer bestimmten Wahrscheinlichkeit auf die richtige Antwort (Validität). Der Einsatz von Telefonjokern muss mit Kosten verbunden sein, damit eine Versuchsperson nicht wahllos (und in beliebiger Reihenfolge) alle Joker beansprucht. Würde sie dies tun, so liesse sich aus der Kaufreihenfolge der Joker nicht mehr auf die Anwendung einer bestimmten Informationssuch-Strategie schliessen. Pro Aufgabe darf eine Versuchsperson – unabhängig von ihrem momentanen Kontostand – 0 bis 6 Joker einsetzen. Wird ein gewählter Joker telefonisch erreicht, so erhält die Versuchsperson von diesem

einen Tipp auf genau eine der vier Antwortalternativen. Wird der Joker nicht erreicht – dieser Fall tritt mit der Gegenwahrscheinlichkeit von D ein – erhält die Versuchsperson entsprechend auch keinen Tipp. Die Kosten für das Anwählen des Jokers werden aber in jedem Fall in Rechnung gestellt.

4.3.3 Vier Antwort-Alternativen

Um zu verhindern, dass eine Vp versucht die richtige Antwort zu erraten (weil sie sich die Kosten für den Einsatz von Telefonjokern sparen will), bieten wir pro Frage vier Antwort-Alternativen an, anstatt nur zwei wie in bisher durchgeführten Experimenten zu Urteilsheuristiken. Dadurch sinkt die Ratewahrscheinlichkeit von 0.5 (bei zwei Alternativen) auf 0.25 (bei 4 Alternativen). Reines Raten lohnt sich damit nicht mehr – immer vorausgesetzt, die Versuchsperson kann bei der entsprechenden Aufgabe kein Vorwissen anwenden, um eine oder mehrere der Antwort-Alternativen von vornherein auszuschliessen.

4.3.4 Natürliche Präsentationsform der entscheidungspsychologischen Konzepte

Tabelle 4.1 fasst zusammen, wie die entscheidungs-psychologischen Konzepte 'Cue', 'Diskriminationsrate' und 'Validität' in das Experiment eingebettet sind, um eine möglichst natürliche Präsentationsform der Konzepte zu erreichen.

Tabelle 4.1: Entscheidungspsychologische Konzepte und ihre Abbildung im Experiment

Entscheidungspsychologisches Konzept	Abbildung des entscheidungspsychologischen Konzepts (Einbettung in das Experiment)
Cue	Telefonjoker (kurz: Joker)
Cue-Kauf	Einsatz eines Telefonjokers (d.h. Telefonanruf)
Cue-Information	Tipp eines Telefonjokers (kurz: Joker-Tipp)
Cue-Set	Telefonjoker-Gruppe (kurz: Joker-Gruppe)
Diskriminationsrate (D)	Telefonische Erreichbarkeit eines Jokers
Validität (V)	Zuverlässigkeit (eines Jokers) bezüglich des Wissensgebiets der momentan angezeigten Aufgabe

4.3.5 Feedback und Motivation der Versuchspersonen

Ein wichtiger Punkt bei der Festlegung eines Versuchsdesigns zum probabilistischen Entscheiden besteht in der Frage, ob die Versuchspersonen eine Rückmeldung über die Korrektheit der getroffenen Entscheidung erhalten sollen und falls ja wann. In der Regel ist, um eine Vergleichbarkeit der Experimentaldurchgänge über das gesamte Experiment zu gewährleisten, Lernen

aufgrund von Feedback unerwünscht und deswegen zu verhindern. Ohne Feedback ergeben sich aber motivationale Probleme. Das vorliegende Design löst dieses Dilemma, indem in jedem Durchgang wieder sechs neue Telefonjoker präsentiert werden. Ausserdem gibt es jeweils eine neue Wissensfrage, so dass die Versuchspersonen hinsichtlich der folgenden Durchgänge nicht von den bisherigen Durchgängen profitieren können.

Deswegen können die Versuchspersonen unmittelbar nach jeder definitiven Entscheidung für eine der Antwort-Alternativen eine Rückmeldung über die Korrektheit ihrer Entscheidung erhalten. Dieses Feedback spielt motivational eine wichtige Rolle, weil Neugierde auf die richtige Lösung geweckt und die Leistungsmotivation angesprochen wird. Zudem haben die Versuchspersonen dank dem Feedback die Möglichkeit, etwas dazuzulernen und damit ein klein wenig schlauer aus dem Experiment herauszugehen sie hineingekommen sind.

Wichtigster Motivationsfaktor für eine seriöse Einstellung zur Erbringung einer möglichst guten Leistung ist allerdings die Ausschüttung eines kleinen Gewinns für jede richtig beantwortete Frage. Der über alle Aufgaben summierte Spielsaldo wird am Ende des Experiments in einen realen Franken-Betrag umgerechnet und sofort bar ausbezahlt. Damit ist also ein leistungsabhängiger, realer monetärer Gewinn sichergestellt.

Wie in derartigen Experimenten üblich, absolvieren die Versuchspersonen zu Beginn des Experiments einen kurzen Probedurchgang. Die Anrufe der virtuellen Telefonjoker werden dabei akustisch simuliert, um die Versuchspersonen möglichst intensiv und realitätsnah auf die im Experiment vorliegende Spielsituation einzustimmen. Diese akustische S(t)imulation kann als ein weiterer motivationaler Aspekt des Spiels betrachtet werden.

Tabelle 4.2: Mögliche Konfidenzangaben

Konfidenzangaben	
völlig geraten	: 25% (Ratewahrscheinlichkeit)
ziemlich unsicher	: 26% - 60%
eher sicher	: 61% - 90%
ziemlich sicher	: 91% - 99%
absolut sicher	: 100%

4.3.6 Konfidenz-Rating

Zu einigen der Wissensfragen kann die Versuchsperson möglicherweise Vorwissen bei der Beantwortung einsetzen. Je nachdem wie ausgeprägt dieses Vorwissen ist, wird das ihr Informationssuch-Verhalten gegenüber dem „Normalfall“ der Unwissenheit stark verändern. Um diese Fälle bei der Auswertung berücksichtigen zu können, gibt die Versuchsperson bei jeder Frage ein Konfidenz-Rating ab, bevor sie sich auf Informationssuche machen kann und Telefonjoker anruft. Mit dieser Angabe ihrer Antwort-Sicherheit (Konfidenz) kann nicht nur überprüft wer-

den, ob die Versuchsperson in diesem Fall Vorwissen eingesetzt hat (im Sinne des Wissens um die richtige Antwort oder aber im Sinne einer gewissen Ahnung). Ein zweites Konfidenz-Rating bei der Abgabe der definitiven Antwort ermöglicht auch, den Zugewinn an Sicherheit angesichts der Auskünfte der Telefonjoker zu messen. Für den Zweck dieses Experimentes ist eine fünfstufige Konfidenz-Abfrage (Tabelle 4.2) praktikabel.

4.3.7 Messen der Suchregel in zwei Phasen

Um eine ausreichend genaue Detektion der Suchstrategie-Anwendung zu erreichen, ist das Experiment in zwei Phasen unterteilt: In einer ersten Phase findet eine Eingrenzung von fünf auf zwei Strategien statt. In der zweiten Phase werden dann diese beiden Strategien gegeneinander getestet (Abbildung 4.1). Die beiden Phasen sind grundsätzlich identisch aufgebaut. Aus Sicht der Versuchspersonen unterscheiden sie sich nur in der Kostenstruktur voneinander. Die zweite Phase benötigt einen kleineren Aufgabenumfang als die erste, da mit ihr lediglich zwei Strategien voneinander zu unterscheiden sind.

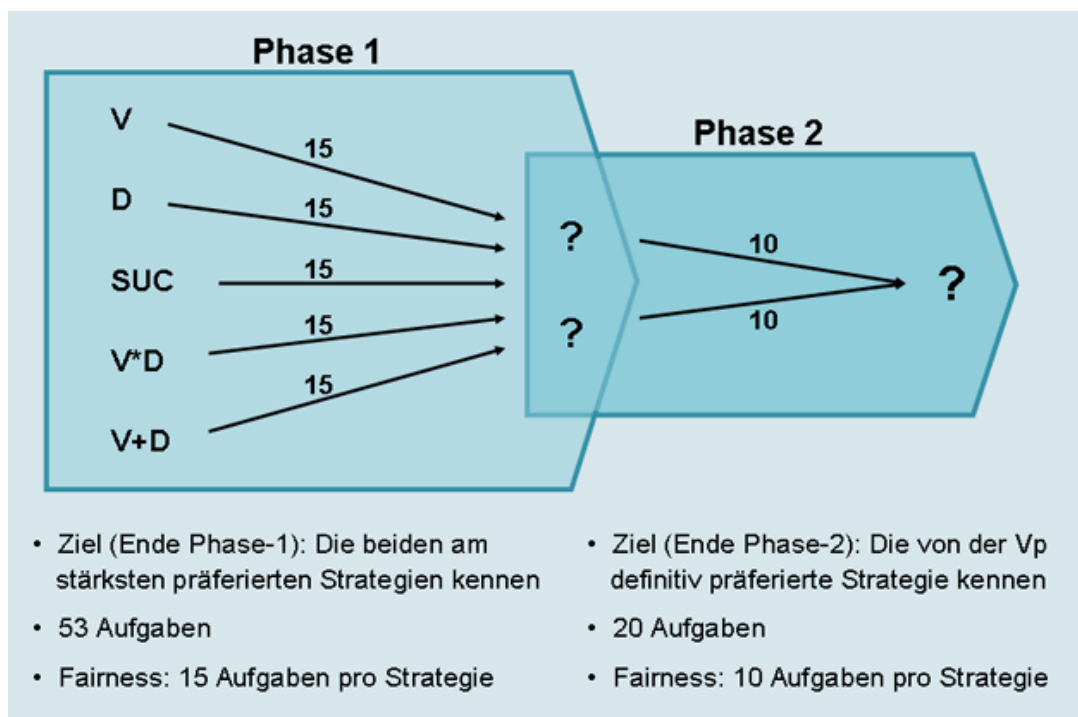


Abbildung 4.1: Die beiden Phasen des Hauptdurchgangs

Der Umfang der einzelnen Phasen muss zwei Kriterien optimieren, die Genauigkeit der Messung und die ökonomische Durchführbarkeit des Experiments. Zur Lösung einer Aufgabe (Durchlesen und Beantworten der Frage, Studium der Antwort-Alternativen, Abgabe von ein oder zwei Konfidenz-Ratings, Auswahl von im Mittel zwei Telefonjokern und Durchlesen des Erläuterungstextes) ist eine mittlere Zeit von ca. 40 Sekunden zu veranschlagen. Die zeitliche Limitierung des gesamten Experiments (inklusive Instruktion, Probedurchgang, Präferenzrangreihen-Bildung und Debriefing) auf maximal 70 bis 90 Minuten (d.h. kalkulatorische 80 Minuten) pro Versuchsperson

son führt zu einer Einschränkung auf ca. 75 präsentierbare Aufgaben: Rechnet man mit Zeiten von durchschnittlich je 5 Minuten für die Instruktion, für den Probedurchgang und die Bildung der Präferenzrangreihe sowie ca. 15 Minuten für das Debriefing, so verbleiben im Mittel noch ca. 50 Minuten für den Hauptdurchgang. Bei ca. 40 Sekunden pro Aufgabe ergibt dies eben eine Begrenzung von etwa 75 Aufgaben pro Versuchsperson.

Für eine ausreichend präzise Detektion der zwei am häufigsten verwendeten Suchstrategien (innerhalb der 5 untersuchten Suchstrategien) sind mindestens 15 Aufgaben pro überprüfter Strategie nötig. Es ist nicht möglich, mit den ca. zwei angeschauten Jokern einer Aufgabe alle fünf Strategien voneinander zu unterscheiden, aber durch eine spezielle Konstruktion der Cue-Sets lässt sich zumindest erreichen, im Mittel mehr als nur eine Strategie gegen die vier anderen klar zu testen. (und zwar bereits mit dem ersten oder spätestens dem zweiten Cue-Kauf). Dies führt dazu, dass für Phase 1 anstelle von $5 \cdot 15 = 75$ Aufgaben lediglich 45 Aufgaben zur eigentlichen Strategie-Detektion benötigt werden. Allerdings sind fünf zusätzliche Aufgaben hinzuzufügen, die nicht zur Strategie-Detektion sondern zur Unterbringung spezieller Cue-Sets dienen (um Validitäten und Diskriminationsraten über das gesamte Experiment ausbalancieren zu können). Schliesslich enthält Phase 1 noch weitere drei Aufgaben, in denen jeweils nur ein einziger Telefonjoker eine Zuverlässigkeit von mehr als 25 % (d.h. grösser als die Ratewahrscheinlichkeit) aufweist. Damit lässt sich testen, ob eine Versuchsperson das Konzept der Validität wirklich versteht (da Joker mit einer Wahrscheinlichkeit von weniger als 25 % ja Tipps abgeben, die mit einer grösseren Wahrscheinlichkeit falsch als richtig sind).

Insgesamt benötigt Phase 1 somit 53 Aufgaben. Somit verbleiben für Phase 2 noch ca. 20 Aufgaben zur Entscheidung, welche der beiden in Phase 1 am häufigsten angewendeten Strategien denn nun tatsächlich die von einer Versuchsperson präferierte Strategie darstellt. Dazu stehen pro Strategie 10 Aufgaben zur Verfügung, die diese eindeutig von jeder beliebigen anderen Strategie unterscheidet. Phase 2 kombiniert zwei dieser Sets und macht damit bei jeder Strategie die Vorhersage, dass diese zehnmal detektiert wird und dass die Gegenstrategie nie detektiert wird (denn das Verwenden einer bestimmten Strategie widerspricht in allen zehn Fällen des Gegenstrategie-Sets der Anwendung dieser Gegenstrategie). Phase 2 kann also mit 20 Aufgaben die beiden noch in Frage kommenden Strategien sicher voneinander unterscheiden, so dass das Experiment in der Summe mit 73 Durchgängen auskommt.

4.3.8 Kostenstruktur (Punkte vs. CHF-Beträge)

Das Verhältnis zwischen Kosten und Gewinn soll so gestaltet sein, dass es im Grundsatz favorabel für One-Reason-Decision-Making (und insbesondere die TTB-Heuristik) ist. Dies bedeutet, dass die Joker auf keinen Fall zu wenig kosten dürfen. Wären die Joker gratis oder nahezu gratis, würden die Versuchspersonen wahllos alle Joker einsetzen, wobei die Reihenfolge der Joker-Einsätze dann keine Messung im Sinne von Suchstrategien mehr erlauben würde. Deshalb müssen die Kosten für die einzelnen Telefonjoker so hoch sein, dass eine Versuchsperson gezielt nur diejenigen Joker einsetzt, die sie für die besten hält. Sind die Joker allerdings zu teuer, so setzen die Versuchspersonen nur sehr selten oder vielleicht gar nie welche ein. Es ist also wichtig, ein geeignetes Verhältnis zwischen Kosten pro Joker und möglichem Spielgewinn pro Aufgabe zu

etablieren. Die Erfahrung aus früheren Experimenten zeigt, dass sich ein Kosten/Gewinn-Verhältnis von 1:10 bewährt.

4.3.9 Ablauf des Experiments im Überblick

Nach der Umsetzung der o.g. Anforderungen sieht der grundsätzliche Ablauf des Experiments folgendermassen aus: Zu Beginn wird eine Versuchsperson mit der ihr zugeteilten Funktion (Kandidat bzw. Kandidatin in einer Quizshow) und der zu erledigenden Aufgabe (korrekte Beantwortung von schwierigen Wissensfragen) vertraut gemacht. Die Spielanleitung wird in elektronischer Form präsentiert und kann während des Experiments jederzeit (durch Drücken der Taste F1) wieder angezeigt werden. Nach dem Durchlesen der Spielanleitung absolviert die Versuchsperson einen kurzen, drei Aufgaben umfassenden Probedurchgang. Anhand dessen wird ihr der Ablauf der Aufgaben erklärt und ausserdem wird sie dadurch auf die Spielsituation eingestimmt. Nach Abschluss des Probedurchgangs folgt der wichtigste Teil des Experiments: Der weiter oben beschriebene, aus zwei Phasen bestehende Hauptdurchgang zur Detektion der präferierten Informationssuch-Strategie. Im Anschluss an den Hauptdurchgang muss die Versuchsperson, ebenfalls am Bildschirm, eine Cue-Präferenzrangreihe aus 10 fest vorgegebenen Cues bilden. Diese Rangreihe soll uns zusätzliche Hinweise darauf liefern, ob die Versuchsperson grundsätzlich die Validität oder die Diskriminationsrate eines Cues stärker gewichtet bzw. auf welche Weise sie diese beiden Werte miteinander zu einem kombinierten Gütemass verrechnet. Anschliessend gibt die Versuchsperson am Bildschirm gleich selbst ihre persönlichen Daten (Name, Adresse, Beruf, etc.) sowie optional ein kurzes Feedback zum Experiment ein. Danach erfolgt die Anzeige des erreichten Spielstands sowie (auf Verlangen) die Anzeige des Highscores (Top-5).

4.4 Der Versuchsablauf

4.4.1 Übersicht über die einzelnen Schritte des durchgeführten Quizshow-Experiments

Der Ablauf des Experiments umfasste die folgenden Schritte (siehe hierzu auch Abbildung 4.2):

1. Login (Eingabe von: Name der Versuchspersonen / Name des Versuchsleiters)
2. Begrüssungstext lesen & Durchgang wählen
3. Durchlesen der Spielanleitung
4. Absolvierung des Probedurchgangs
5. Absolvierung des Hauptdurchgangs:
 - a. Absolvierung von Phase 1 ("Hypothesen-Generierung")
 - b. Anzeige des in Phase 1 erreichten Spielstands
 - c. Absolvierung von Phase 2 ("Hypothesenüberprüfung")
 - d. Anzeige des in Phase 2 erreichten Spielstands
6. Bildung einer Präferenzrangreihe (10 Cues)
7. Erfassung der Versuchspersonen-Daten (Name, Adresse, Beruf, etc. + Feedback)
8. Anzeige des Schlussspielstands (Spielgewinn)
9. optional: Anzeige des Highscores (Top-5)
10. Ausbezahlung des erspielten Gewinns

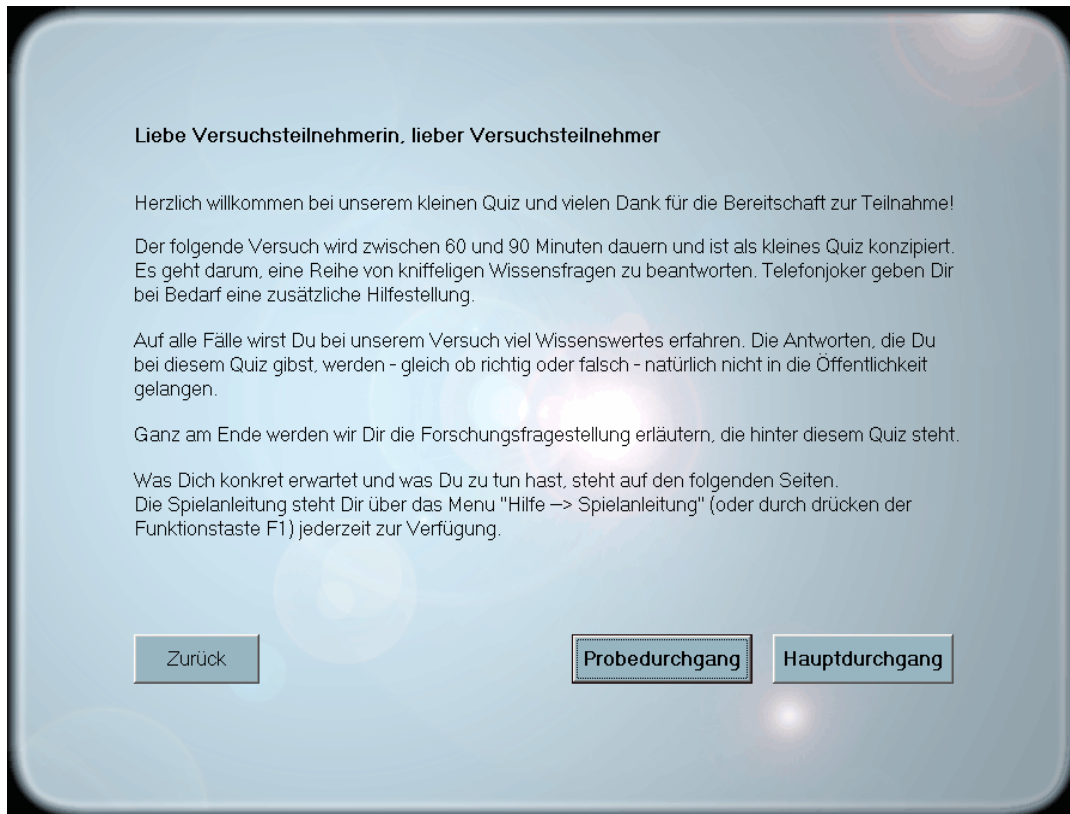


Abbildung 4.4: Begrüßungstext und Auswahl des Versuchsdurchgangs

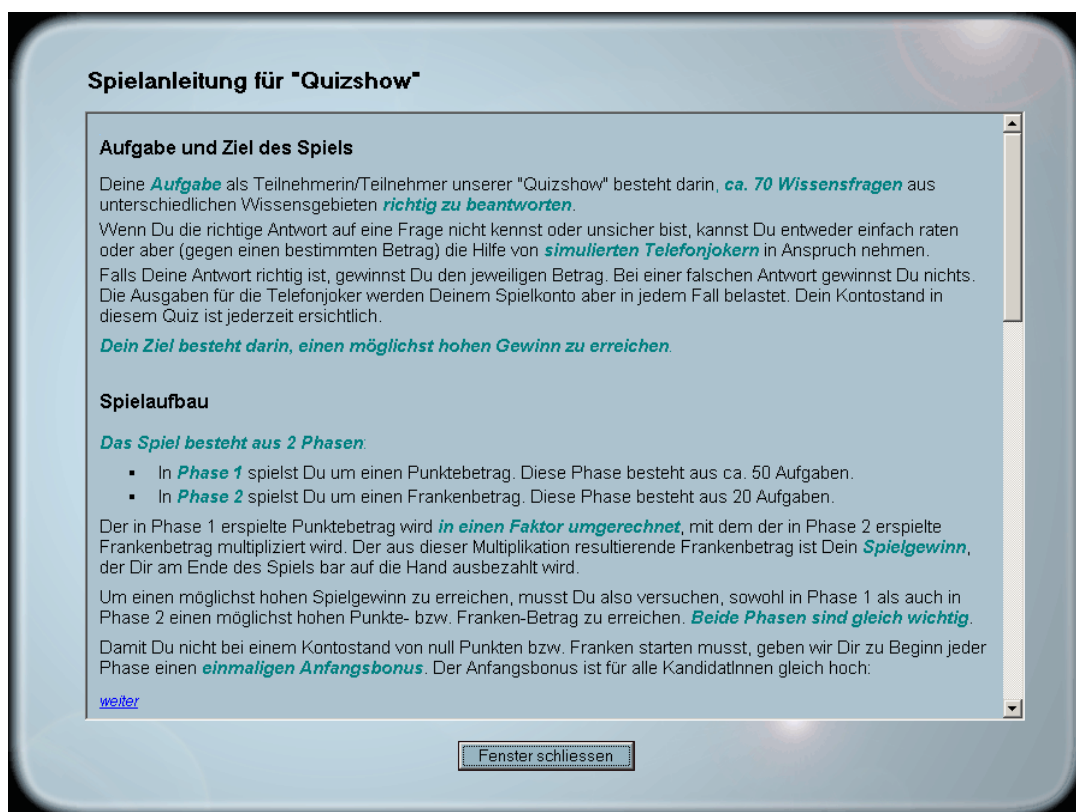


Abbildung 4.5: Die erste Seite der Spielanleitung im Browserfenster

Aufgabe und Ziel des Spiels

Deine **Aufgabe** als Teilnehmerin/Teilnehmer unserer "Quizshow" besteht darin, **ca. 70 Wissensfragen** aus unterschiedlichen Wissensgebieten **richtig zu beantworten**.

Wenn Du die richtige Antwort auf eine Frage nicht kennst oder unsicher bist, kannst Du entweder einfach raten oder aber (gegen einen bestimmten Betrag) die Hilfe von **simulierten Telefonjokern** in Anspruch nehmen.

Falls Deine Antwort richtig ist, gewinnst Du den jeweiligen Betrag. Bei einer falschen Antwort gewinnst Du nichts. Die Ausgaben für die Telefonjoker werden Deinem Spielkonto aber in jedem Fall belastet. Dein Kontostand in diesem Quiz ist jederzeit ersichtlich.

Dein Ziel besteht darin, einen möglichst hohen Gewinn zu erreichen.

Spiel Aufbau

Das Spiel besteht aus 2 Phasen:

In **Phase 1** spielst Du um einen Punktebetrag. Diese Phase besteht aus ca. 50 Aufgaben.

In **Phase 2** spielst Du um einen Frankenbetrag. Diese Phase besteht aus 20 Aufgaben.

Der in Phase 1 erspielte Punktebetrag wird **in einen Faktor umgerechnet**, mit dem der in Phase 2 erspielte Frankenbetrag multipliziert wird. Der aus dieser Multiplikation resultierende Frankenbetrag ist Dein **Spielgewinn**, der Dir am Ende des Spiels bar auf die Hand ausbezahlt wird.

Um einen möglichst hohen Spielgewinn zu erreichen, musst Du also versuchen, sowohl in Phase 1 als auch in Phase 2 einen möglichst hohen Punkte- bzw. Franken-Betrag zu erreichen. **Beide Phasen sind gleich wichtig.**

Damit Du nicht bei einem Kontostand von null Punkten bzw. Franken starten musst, geben wir Dir zu Beginn jeder Phase einen **einmaligen Anfangsbonus**. Der Anfangsbonus ist für alle KandidatInnen gleich hoch:

Anfangsbonus Phase 1: 1'000 Punkte

Anfangsbonus Phase 2: CHF 2.-

Für die Beantwortung der Fragen gibt es **keine Zeitbeschränkung**. Bedenke aber, dass es viele Fragen zu beantworten gibt. Gehe deshalb jeweils relativ zügig weiter zur nächsten Frage.

Schritte bei der Beantwortung einer Quizfrage

Die Beantwortung einer einzelnen Quizfrage beinhaltet **mehrere Schritte**:

1. Als erstes musst Du jeweils die gestellte **Frage beantworten**, indem Du eine der 4 vorgegebenen Antwortalternativen auswählst (Genau eine der 4 Antwortalternativen ist die richtige!)
2. Danach möchten wir von Dir wissen, wie sicher Du Dir bei Deiner ersten Antwort bist. Dazu musst Du eine **Sicherheitsangabe auswählen**. Es stehen 5 Sicherheitsangaben zur Auswahl.
3. Nun kannst Du (wenn Du willst) bis zu maximal 6 **Telefonjoker einsetzen**.
4. (Dies gilt für jede Aufgabe; d.h. Du kannst in jeder Aufgabe aufs Neue bis zu 6 Telefonjoker einsetzen)
5. Nach dem Einsatz eines oder mehrerer Joker musst Du **die gestellte Frage nochmals beantworten**.
6. Zum Schluss **gibst Du nochmals Deine (ggf. revidierte) Antwort-Sicherheit an**.
7. Daraufhin erfährst Du, zusammen mit einer kurzen Erläuterung, welches die richtige Antwort ist. Hoffentlich ist es diejenige, die du gewählt hast!
8. Und weiter geht's zur nächsten Frage.

Wir spielen diese Schritte gleich einmal in der Praxis durch.

Zuvor solltest Du aber unbedingt noch etwas über die Telefonjoker wissen:

Der Einsatz eines Telefonjokers **kostet** Dich einen bestimmten Punkte- bzw. Franken-Betrag. Der Betrag ist jeweils angegeben. Er ist in allen Aufgaben und für alle Joker gleich hoch.

Wenn Du die Hilfe von Telefonjokern in Anspruch nehmen willst, erscheint jeweils eine Auswahlliste mit 6 Telefonjokern. Jeder Joker hat 2 Kennwerte: Seine **"Erreichbarkeit"** und die **"Zuverlässigkeit"** seines Tipps.

Die **Erreichbarkeit** gibt an, mit welcher Wahrscheinlichkeit (ein Wert zwischen 0% und 100%) der betreffende Joker überhaupt telefonisch erreichbar ist.

Die **Zuverlässigkeit** (eine Zahl zwischen 25% und 100%) gibt an, wie zuverlässig der Tipp des betreffenden Jokers ist, bzw. wie sicher oder glaubhaft er ist.

Ist der Tipp absolut sicher, beträgt die Wahrscheinlichkeit 100%, ist er nichts als geraten, beträgt er **bei 4 Antwortalternativen** genau **25%**.

Die Telefonjoker und ihre Kennwerte wechseln von Frage zu Frage. Mit anderen Worten: Im ganzen

Abbildung 4.6: Spielanleitung im Wortlaut

(2) *Begrüßungstext und Auswahl des Versuchs-Durchgangs.* Der genaue Wortlaut des Begrüßungstexts kann Abbildung 4.4 entnommen werden. Auf derselben Bildschirmseite (am Ende des Begrüßungstexts) befand sich die aus zwei Schaltflächen bestehende Durchgangs-Auswahl ("Probedurchgang" / "Hauptdurchgang").

(3) *Spielanleitung.* Die Auswahl des Probedurchgangs (im Hauptmenü) führte zunächst zur Anzeige der Spielanleitung. Diese musste vor der Absolvierung der Probeaufgaben durchgelesen werden. Während des Hauptdurchgangs konnte sie durch Knopfdruck (Taste F1) jederzeit wieder eingeblendet werden. Da sie in einem integrierten Browserfenster angezeigt wurde (Abbildung 4.5), musste sie als HTML-Dokument vorliegen. Sie konnte somit bequem als Worddokument erstellt und dann als HTML-Dokument gespeichert werden. Zur ihrer vollständigen Darstellung auf dem Bildschirm wurden insgesamt 3 Seiten im Browserfenster benötigt. Ihr genauer Wortlaut kann Abbildung 4.6 entnommen werden.

(4) *Probedurchgang.* Der Probedurchgang war für die Versuchspersonen als solcher deklariert. Er umfasste für alle Versuchspersonen dieselben drei Aufgaben (Fragen) in derselben Reihenfolge. In diesen drei Fragen wurde jeweils um Punkte gespielt. Die erspielten Punkte wurden mit Beendigung des Probedurchgangs verworfen und hatten damit keinerlei Bedeutung mehr für den weiteren Verlauf des Experiments. Der Probedurchgang konnte bei Bedarf beliebig oft wiederholt werden.

Frage

Wer erfand 1800 die erste Batterie?

James Watt A

Alessandro Volta B

André Marie Ampère C

William Henry D

OK

Kontostand (Punkte) - Probedurchgang

Einnahme bei richtiger Antwort: 1'000

Bisheriger Kontostand: 1'000

Ausgaben bei aktueller Frage: 200

Einnahme bei aktueller Frage: ?

Resultierender Gewinn / Verlust: ?

Neuer Kontostand: ?

Antworten Total: 0 (falsche: 0 / richtige: 0)

1 2 3 4 5 6 7

Nun kannst Du Telefonjoker einsetzen

Telefonjoker

Nr	Kosten	Erreichbarkeit	Zuverlässigkeit	Tipp
1	100 Pkte	93 %	35 %	
2	100 Pkte	57 %	36 %	
3	100 Pkte	19 %	41 %	
4	100 Pkte	68 %	90 %	
5	100 Pkte	51 %	70 %	B
6	100 Pkte	37 %	78 %	-

Bisherige Ausgaben: 200 Pkte

Mögliche Einnahme: 1'000 Pkte

Fertig mit kaufen

Antwort (Erläuterung)

Abbildung 4.7: Probedurchgang, Frage 1

Frage

Wo erschien 1605 die erste Zeitung?

Strassburg A

Leipzig B

Neu Amsterdam C

London D

Wie sicher bist Du? 61% - 90% (eher sicher) OK

Kontostand (Punkte) - Phase 1

Einnahme bei richtiger Antwort: 1'000

Bisheriger Kontostand: 8'200

Ausgaben bei aktueller Frage: 300

Einnahme bei aktueller Frage: ?

Resultierender Gewinn / Verlust: ?

Neuer Kontostand: ?

Antworten Total: 16 (falsche: 5 / richtige: 11)

1 2 3 4 5 6 7

Gib nochmals an, wie sicher Du Dir bist

Telefonjoker

Nr	Kosten	Erreichbarkeit	Zuverlässigkeit	Tipp
1	100 Pkte	19 %	67 %	A
2	100 Pkte	58 %	50 %	
3	100 Pkte	71 %	54 %	
4	100 Pkte	93 %	37 %	
5	100 Pkte	46 %	93 %	—
6	100 Pkte	6 %	89 %	—

Bisherige Ausgaben: 300 Pkte

Mögliche Einnahme: 1'000 Pkte

Fertig mit kaufen

Antwort (Erläuterung)

Abbildung 4.8: Hauptdurchgang, Phase 1

Um die Versuchspersonen in möglichst plastischer Weise auf die Spielsituation einzustimmen, wurde der Anruf eines Telefonjokers im Probedurchgang akustisch simuliert. Zu hören war jeweils der Verbindungsaufbau (Ruftön) sowie die Antwort des Jokers, falls dieser erreicht wurde (Beispiel: "Ja hallo! Hier ist der Thomas. Ich tippe auf Antwort B. Viel Glück. Tschüss."). Falls der Joker nicht erreicht wurde, war entweder a) seine Telefonbeantworter-Ansage zu hören oder b) der Verbindungsaufbau wurde nach 5 Rufzeichen erfolglos wieder abgebrochen oder c) ein Störungszeichen ertönte (schnelles Tuten). Mit Ausnahme der akustischen Simulation der Telefonjoker war der Probedurchgang vom look-and-feel her gesehen identisch mit Phase 1 des Hauptdurchgangs (Abbildung 4.7; vgl. Abbildung 4.8).

(5) *Hauptdurchgang*. Der Hauptdurchgang umfasste 73 Aufgaben und war in 2 Phasen unterteilt: In Phase 1 (53 Aufgaben) spielten die Versuchspersonen jeweils um Punkte (Abbildungen 4.8 und 4.9) und in Phase 2 (20 Aufgaben) jeweils um Franken-Beträge (Abbildungen 4.10 und 4.11). Durch die Umstellung auf Franken-Beträge in Phase 2 sollten die Versuchspersonen nochmals für die Bedeutung ihrer Kaufentscheidungen sensibilisiert werden, d.h. auf die Wichtigkeit möglichst den "besten" Cue zuerst zu kaufen, um die Ausgaben möglichst tief zu halten und gleichzeitig möglichst oft einen Gewinn einzustreichen. Die Aufteilung in Punkte (Phase 1) und Franken-Beträge (Phase 2) ermöglichte ausserdem auf einfache Weise eine auch für die Versuchspersonen jederzeit leicht nachvollziehbare Berechnung des resultierenden, am Ende des Versuchs bar ausbezahlten Spielgewinns. Der Hauptdurchgang konnte sowohl in Phase 1 als

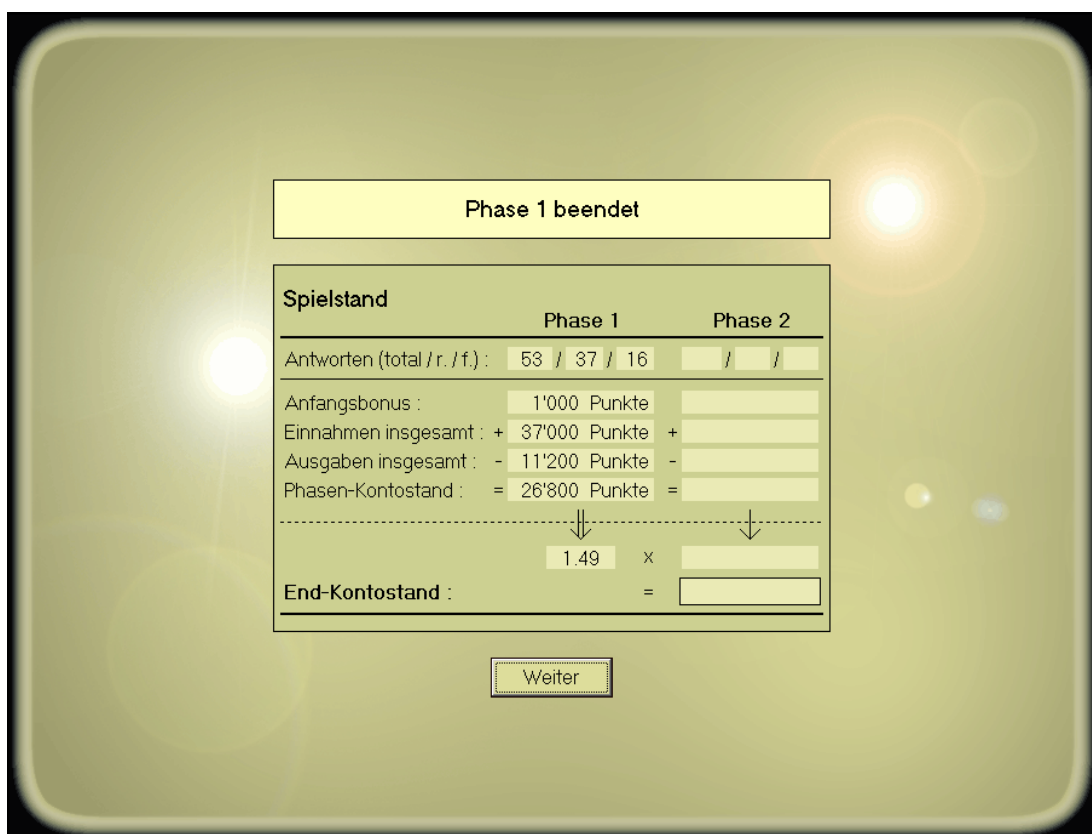
auch in Phase 2 jederzeit unterbrochen und zu einem späteren Zeitpunkt an derselben Stelle mit einer neuen Frage fortgesetzt werden.

(5a) *Absolvierung von Phase 1.* Abbildung 4.8 illustriert den Bildschirmaufbau von Phase 1. Sie umfasste 53 Aufgaben und diente der "Hypothesengenerierung" auf individueller Ebene, d.h. der Evaluation der zwei am stärksten präferierten Informationssuch-Strategien einer Versuchsperson. Unmittelbar nach der Lösung der letzten Aufgabe dieser Phase analysierte das Programm, welche beiden Strategien die Versuchsperson am häufigsten angewendet hatte, um sie in Phase 2 gegeneinander testen zu können. Primäres Auswahlkriterium für die beiden Strategien war ihre Anwendungshäufigkeit.

(5b) *Spielstand nach Phase 1.* Nach Abschluss von Phase 1 folgte die Anzeige des in dieser Phase erreichten Spielstands (Abbildung 4.9).

(5c) *Absolvierung von Phase 2.* Phase 2 (Abbildung 4.10) umfasste 20 Aufgaben. Diese Phase diente der "Hypothesen-Überprüfung". In ihr wurden die beiden von der Vp in Phase 1 am häufigsten verwendeten Strategien noch einmal exklusiv gegeneinander getestet, um zu ermitteln, welche der beiden Strategien sie effektiv präferiert.

(5d) *Spielstand nach Phase 2.* Nach Abschluss von Phase 2 wurde wiederum der erreichte Spielstand angezeigt. Ersichtlich war sowohl der in Phase 1 als auch der in Phase 2 erzielte Stand. Zudem wurde der resultierende Spielgewinn angezeigt (Abbildung 4.11).



Phase 1 beendet		
Spielstand		
	Phase 1	Phase 2
Antworten (total / r. / f.) :	53 / 37 / 16	/ / /
Anfangsbonus :	1'000 Punkte	
Einnahmen insgesamt : +	37'000 Punkte +	
Ausgaben insgesamt : -	11'200 Punkte -	
Phasen-Kontostand : =	26'800 Punkte =	
	↓	↓
	1.49 x	
End-Kontostand :	=	

Weiter

Abbildung 4.9: Spielstandanzeige nach Abschluss von Phase 1

Frage

Welches ost-west-gerichtete Urmeer bildete sich im Paläozoikum?

Okeanos

Tethys

Pontos

Atlantik

A

B

C

D

Wie sicher bist Du? 61% - 90% (eher sicher) OK

Kontostand (CHF) - Phase 2

Einnahme bei richtiger Antwort:	1.00
Bisheriger Kontostand:	10.60
Ausgaben bei aktueller Frage:	0.20
Einnahme bei aktueller Frage:	?
Resultierender Gewinn / Verlust:	?
Neuer Kontostand:	?

Antworten Total: 17 (falsche: 4 / richtige: 13)

1
2
3
4
5
6
7

Gib nochmals an, wie sicher Du Dir bist

Telefonjoker

Nr	Kosten	Erreichbarkeit	Zuverlässigkeit	Tipp
1	CHF 0.10	15 %	53 %	
2	CHF 0.10	63 %	93 %	—
3	CHF 0.10	30 %	88 %	B
4	CHF 0.10	93 %	40 %	
5	CHF 0.10	20 %	71 %	
6	CHF 0.10	66 %	63 %	

Bisherige Ausgaben : CHF 0.20

Mögliche Einnahme : CHF 1.00

Fertig mit kaufen

Antwort (Erläuterung)

Abbildung 4.10: Hauptdurchgang, Phase 2

Phase 2 beendet

Spielstand

	Phase 1	Phase 2
Antworten (total / r. / f.):	53 / 37 / 16	20 / 14 / 6
Anfangsbonus:	1'000 Punkte	2.00 CHF
Einnahmen insgesamt: +	37'000 Punkte	+ 14.00 CHF
Ausgaben insgesamt: -	11'200 Punkte	- 3.80 CHF
Phasen-Kontostand: =	26'800 Punkte	= 12.20 CHF
	↓	↓
	1.49	x 12.20 CHF
End-Kontostand:	=	CHF 18.18

Weiter

Abbildung 4.11: Spielstand-Anzeige nach Abschluss von Phase 2

Abbildung 4.12: Präferenzrangreihen-Bildung mit den 10 Cues

Abbildung 4.13: Erfassung persönlicher Angaben



Abbildung 4.14: Spielstand-Anzeige am Ende des Experiments



Abbildung 4.15: Anzeige der Top-5-Highscore-Liste (optional)

(6) *Bildung einer Präferenzrangreihe mit 10 Cues.* Im Anschluss an den Hauptdurchgang mussten die Versuchspersonen mit 10 fest vorgegebenen Cues eine Präferenzrangreihe bilden (Abbildung 4.12). Jeder Versuchsperson wurden die gleichen Cues vorgelegt. Die erstellte Präferenzrangreihe sollte einen Hinweis darauf liefern, ob sie bei der Informationssuche die Validität oder die Diskriminationsrate stärker gewichtet bzw. ob sie die beiden Masse zur Bewertung der Cue-Güte miteinander verrechnet.

(7) *Angaben der Versuchsperson.* Die Erfassung der Versuchspersonen-Daten (Name, Adresse, Beruf, etc.) mittels On-Screen-Formular (Abbildung 4.13) hatte sich bereits in früheren Experimenten bewährt. Deshalb liessen wir die Versuchspersonen ihre persönlichen Daten gleich selbst eingeben. Im selben Formular konnten sie uns auch ihr Feedback zum Experiment mitteilen.

(8) *Anzeige des Schlussspielstands.* Nachdem eine Versuchsperson ihre persönlichen Daten sowie allenfalls ein Feedback zum Experiment eingegeben hatte, wurde der von ihr erreichte End-Spielstand angezeigt (Abbildung 4.14). Ersichtlich waren sowohl die in den beiden Phasen erreichten Spielstände als auch der daraus errechnete Spielgewinn.

(9) *Anzeige des Highscores (optional).* Auf ausdrücklichen Wunsch hin konnten sich die Versuchspersonen den Top-5-Highscore ansehen (Abbildung 4.15). Der Highscore erschien nicht automatisch, sondern musste vom Versuchsleiter mittels einer speziellen Tastenkombination aufgerufen werden. Wir wollten damit verhindern, dass Versuchspersonen, die nur einen kleinen Gewinn erzielten, durch die Bekanntgabe besserer Resultate anderer Personen frustriert werden. Es war jedoch so, dass ausnahmslos alle Versuchspersonen sich den Highscore ansehen wollten.

(10) *Auszahlung des erspielten Gewinns.* Die Auszahlung des erspielten Gewinns erfolgte unmittelbar nach dem Ende des Experiments. Der Gewinn wurde auf den nächst-höheren ganzzahligen Frankenbetrag aufgerundet und bar ausbezahlt. Der Empfang des ausbezahlten Betrages musste von den Versuchspersonen schriftlich quittiert werden.

4.5 Ablauf einer Aufgabe

Die Bearbeitung einer Aufgabe verlief in 7 oder 4 Schritten, abhängig davon, ob Telefonjoker eingesetzt wurden oder nicht (Tabellen 4.3 und 4.4).

Tabelle 4.3: Aufgaben-Schritte (mit Joker-Einsatz)

Schritt	Beschreibung	Abbildungen
1	Frage lesen und eine der 4 Antwort-Alternativen auswählen	4.16 / 4.17
2	Antwort-Sicherheit angeben & mit 'OK' bestätigen	4.18 / 4.19
3	Entscheiden ob Telefonjoker eingesetzt werden sollen. Antwort: Ja	4.20
4	0 bis 6 Telefonjoker einsetzen & Ende der "Kaufphase" bestätigen	4.21 – 4.23
5	Nochmals eine der 4 Antwort-Alternativen auswählen	4.24
6	Nochmals Antwort-Sicherheit angeben & mit 'OK' bestätigen	4.25
7	Lösung und Erläuterungstext ansehen & weiter zur nächsten Frage	4.26

Tabelle 4.4: Aufgaben-Schritte (ohne Joker-Einsatz)

Schritt	Beschreibung	Abbildungen
1	Frage lesen und eine der 4 Antwort-Alternativen auswählen	4.16 / 4.17
2	Antwort-Sicherheit angeben & mit 'OK' bestätigen	4.18 / 4.19
3	Entscheiden ob Telefonjoker eingesetzt werden sollen. Antwort: Nein	4.20
4	Lösung und Erläuterungstext ansehen & weiter zur nächsten Frage	4.26

Die Auswahl der Wissensfragen erfolgte, ausser im Probedurchgang, zufällig – im Gegensatz zur Auswahl der Telefonjoker. Die Telefonjoker-Gruppen (Cue-Sets) wurden vorgängig zusammengestellt und waren für alle Versuchspersonen fix (wenn auch in jeweils unterschiedlicher Reihenfolge). Die Reihenfolgen waren gegeneinander ausbalanciert, so dass die Cue-Sets für keine zwei Versuchspersonen in derselben Reihenfolge präsentiert wurden.

The screenshot displays a game interface with four main sections:

- Frage (Question):** Contains the question "Was ist ein Belemnit?" and four answer options: "türkischer Fusssoldat" (A), "Kopffüsser" (B), "Tintenfisch" (C), and "Wüstenbewohner" (D). An "OK" button is located below the options.
- Kontostand (CHF) - Phase 2 (Scoreboard):** Displays the current score and game statistics:
 - Einnahme bei richtiger Antwort: 1.00
 - Bisheriger Kontostand: 11.60
 - Ausgaben bei aktueller Frage: 0.00
 - Einnahme bei aktueller Frage: ?
 - Resultierender Gewinn / Verlust: ?
 - Neuer Kontostand: ?
 - Antworten Total: 18 (falsche: 4 / richtige: 14)
- Telefonjoker:** A large empty box for the phone joker feature.
- Antwort (Erläuterung) (Answer/Explanation):** A large empty box for the answer and explanation.

At the bottom, there is a row of numbered buttons (1-7) and a text prompt: "Bitte beantworte die Frage (links oben)".

Abbildung 4.16: Schritt 1a (Wissensfrage lesen)

Frage

Was ist ein Belemnit?

türkischer Fusssoldat A

Kopffüsser B

Tintenfisch C

Wüstenbewohner D

Wie sicher bist Du? OK

Kontostand (CHF) - Phase 2

Einnahme bei richtiger Antwort: 1.00

Bisheriger Kontostand: 11.60

Ausgaben bei aktueller Frage: 0.00

Einnahme bei aktueller Frage: ?

Resultierender Gewinn / Verlust: ?

Neuer Kontostand: ?

Antworten Total: 18 (falsche: 4 / richtige: 14)

1 2 3 4 5 6 7

Gib an wie sicher Du bist. Danach drücke <OK>

Telefonjoker

Antwort (Erläuterung)

Abbildung 4.17: Schritt 1b (Antwort auswählen)

Frage

Was ist ein Belemnit?

türkischer Fusssoldat A

Kopffüsser B

Tintenfisch C

Wüstenbewohner D

Wie sicher bist Du? OK

Bitte auswählen!

25% (völlig geraten)

26% - 60% (ziemlich unsicher)

61% - 90% (eher sicher)

91% - 99% (ziemlich sicher)

100% (absolut sicher)

Kontostand (CHF) - Phase 2

Einnahme bei richtiger Antwort: 1.00

Bisheriger Kontostand: 11.60

Ausgaben bei aktueller Frage: 0.00

Einnahme bei aktueller Frage: ?

Resultierender Gewinn / Verlust: ?

Neuer Kontostand: ?

Antworten Total: 18 (falsche: 4 / richtige: 14)

1 2 3 4 5 6 7

Gib an wie sicher Du bist. Danach drücke <OK>

Antwort (Erläuterung)

Abbildung 4.18: Schritt 2a (Antwort-Sicherheit auswählen)

Frage

Was ist ein Belemnit?

- türkischer Fusssoldat A
- Kopffüsser B
- Tintenfisch** C
- Wüstenbewohner D

Wie sicher bist Du? (völlig geraten)

Kontostand (CHF) - Phase 2

Einnahme bei richtiger Antwort:	1.00
Bisheriger Kontostand:	11.60
Ausgaben bei aktueller Frage:	0.00
Einnahme bei aktueller Frage:	?
Resultierender Gewinn / Verlust:	?
Neuer Kontostand:	?
Antworten Total: 18 (falsche: 4 / richtige: 14)	

1 2 3 4 5 6 7

Gib an wie sicher Du bist. Danach drücke <OK>

Telefonjoker

Antwort (Erläuterung)

Abbildung 4.19: Schritt 2b (Antwort-Sicherheit bestätigen)

Frage

Was ist ein Belemnit?

- türkischer Fusssoldat A
- Kopffüsser B
- Tintenfisch C
- Wüstenbewohner D

Kontostand (CHF) - Phase 2

Einnahme bei richtiger Antwort:	1.00
Bisheriger Kontostand:	11.60
Ausgaben bei aktueller Frage:	0.00
Einnahme bei aktueller Frage:	?
Resultierender Gewinn / Verlust:	?
Neuer Kontostand:	?
Antworten Total: 18 (falsche: 4 / richtige: 14)	

1 2 3 4 5 6 7

Möchtest Du Telefonjoker einsetzen?

Telefonjoker

Du hast nun die Möglichkeit einen oder mehrere Telefonjoker einzusetzen. Möchtest du von dieser Möglichkeit Gebrauch machen?

Antwort (Erläuterung)

Abbildung 4.20: Schritt 3 (Entscheiden über Joker-Einsatz)

Frage

Was ist ein Belemnit?

türkischer Fusssoldat

Kopffüßer

Tintenfisch

Wüstenbewohner

A
B
C
D

Kontostand (CHF) - Phase 2

Einnahme bei richtiger Antwort:	1.00
Bisheriger Kontostand:	11.60
Ausgaben bei aktueller Frage:	0.00
Einnahme bei aktueller Frage:	?
Resultierender Gewinn / Verlust:	?
Neuer Kontostand:	?

Antworten Total: 18 (falsche: 4 / richtige: 14)

1

2

3

4

5

6

7

Nun kannst Du Telefonjoker einsetzen

Telefonjoker

Nr	Kosten	Erreichbarkeit	Zuverlässigkeit	Tipp
1	CHF 0.10	32 %	75 %	<input type="checkbox"/>
2	CHF 0.10	12 %	36 %	<input type="checkbox"/>
3	CHF 0.10	62 %	63 %	<input type="checkbox"/>
4	CHF 0.10	14 %	64 %	<input type="checkbox"/>
5	CHF 0.10	51 %	91 %	<input type="checkbox"/>
6	CHF 0.10	89 %	56 %	<input type="checkbox"/>

Bisherige Ausgaben : CHF 0.00

Mögliche Einnahme : CHF 1.00

Antwort (Erläuterung)

Abbildung 4.21: Schritt 4a (Joker auswählen)

Frage

Was ist ein Belemnit?

türkischer Fusssoldat

Kopffüßer

Tintenfisch

Wüstenbewohner

A
B
C
D

Kontostand (CHF) - Phase 2

Einnahme bei richtiger Antwort:	1.00
Bisheriger Kontostand:	11.60
Ausgaben bei aktueller Frage:	0.10
Einnahme bei aktueller Frage:	?
Resultierender Gewinn / Verlust:	?
Neuer Kontostand:	?

Antworten Total: 18 (falsche: 4 / richtige: 14)

1

2

3

4

5

6

7

Nun kannst Du Telefonjoker einsetzen

Telefonjoker

Nr	Kosten	Erreichbarkeit	Zuverlässigkeit	Tipp
1	CHF 0.10	32 %	75 %	<input type="checkbox"/>
2	CHF 0.10	12 %	36 %	<input type="checkbox"/>
3	CHF 0.10	62 %	63 %	<input type="checkbox"/>
4	CHF 0.10	14 %	64 %	<input type="checkbox"/>
5	CHF 0.10	51 %	91 %	<input checked="" type="checkbox"/>
6	CHF 0.10	89 %	56 %	<input type="checkbox"/>

Bisherige Ausgaben : CHF 0.10

Mögliche Einnahme : CHF 1.00

Antwort (Erläuterung)

Abbildung 4.22: Schritt 4b (Joker einsetzen)

Frage

Was ist ein Belemnit?

türkischer Fussoldat A

Kopffüßer B

Tintenfisch C

Wüstenbewohner D

OK

Kontostand (CHF) - Phase 2

Einnahme bei richtiger Antwort: 1.00

Bisheriger Kontostand: 11.60

Ausgaben bei aktueller Frage: 0.20

Einnahme bei aktueller Frage: ?

Resultierender Gewinn / Verlust: ?

Neuer Kontostand: ?

Antworten Total: 18 (falsche: 4 / richtige: 14)

1 2 3 4 5 6 7

Nun kannst Du Telefonjoker einsetzen

Telefonjoker

Nr	Kosten	Erreichbarkeit	Zuverlässigkeit	Tipp
1	CHF 0.10	32 %	75 %	
2	CHF 0.10	12 %	36 %	
3	CHF 0.10	62 %	63 %	
4	CHF 0.10	14 %	64 %	
5	CHF 0.10	51 %	91 %	-
6	CHF 0.10	89 %	56 %	B

Bisherige Ausgaben: CHF 0.20

Mögliche Einnahme: CHF 1.00

Fertig mit kaufen

Antwort (Erläuterung)

Abbildung 4.23: Schritt 4c (Ggf. weitere Joker wählen/einsetzen)

Frage

Was ist ein Belemnit?

türkischer Fussoldat A

Kopffüßer B

Tintenfisch C

Wüstenbewohner D

Wie sicher bist Du? 26% - 60% (ziemlich unsicher) OK

Kontostand (CHF) - Phase 2

Einnahme bei richtiger Antwort: 1.00

Bisheriger Kontostand: 11.60

Ausgaben bei aktueller Frage: 0.20

Einnahme bei aktueller Frage: ?

Resultierender Gewinn / Verlust: ?

Neuer Kontostand: ?

Antworten Total: 18 (falsche: 4 / richtige: 14)

1 2 3 4 5 6 7

Gib nochmals an, wie sicher Du Dir bist

Telefonjoker

Nr	Kosten	Erreichbarkeit	Zuverlässigkeit	Tipp
1	CHF 0.10	32 %	75 %	
2	CHF 0.10	12 %	36 %	
3	CHF 0.10	62 %	63 %	
4	CHF 0.10	14 %	64 %	
5	CHF 0.10	51 %	91 %	-
6	CHF 0.10	89 %	56 %	B

Bisherige Ausgaben: CHF 0.20

Mögliche Einnahme: CHF 1.00

Fertig mit kaufen

Antwort (Erläuterung)

Abbildung 4.24: Schritt 5 (Nochmals eine der vier Antwort-Alternativen auswählen)

Frage

Was ist ein Belemnit?

türkischer Fussoldat A

Kopffüßer B

Tintenfisch C

Wüstenbewohner D

Wie sicher bist Du? 26% - 60% (ziemlich unsicher) OK

Telefonjoker

Nr	Kosten	Erreichbarkeit	Zuverlässigkeit	Tipp
1	CHF 0.10	32 %	75 %	
2	CHF 0.10	12 %	36 %	
3	CHF 0.10	62 %	63 %	
4	CHF 0.10	14 %	64 %	
5	CHF 0.10	51 %	91 %	—
6	CHF 0.10	89 %	56 %	B

Bisherige Ausgaben : CHF 0.20

Mögliche Einnahme : CHF 1.00

Fertig mit kaufen

Kontostand (CHF) - Phase 2

Einnahme bei richtiger Antwort : 1.00

Bisheriger Kontostand : 11.60

Ausgaben bei aktueller Frage : 0.20

Einnahme bei aktueller Frage : ?

Resultierender Gewinn / Verlust : ?

Neuer Kontostand : ?

Antworten Total: 18 (falsche: 4 / richtige: 14)

1 2 3 4 5 6 7

Gib nochmals an, wie sicher Du Dir bist

Antwort (Erläuterung)

Abbildung 4.25: Schritt 6 (Nochmals Antwort-Sicherheit auswählen und bestätigen)

Frage

Was ist ein Belemnit?

türkischer Fussoldat A

Kopffüßer B

Tintenfisch C

Wüstenbewohner D

Wie sicher bist Du? 26% - 60% (ziemlich unsicher) OK

Telefonjoker

Nr	Kosten	Erreichbarkeit	Zuverlässigkeit	Tipp
1	CHF 0.10	32 %	75 %	
2	CHF 0.10	12 %	36 %	
3	CHF 0.10	62 %	63 %	
4	CHF 0.10	14 %	64 %	
5	CHF 0.10	51 %	91 %	—
6	CHF 0.10	89 %	56 %	B

Kontostand (CHF) - Phase 2

Einnahme bei richtiger Antwort : 1.00

Bisheriger Kontostand : 11.60

Ausgaben bei aktueller Frage : 0.20

Einnahme bei aktueller Frage : 1.00

Resultierender Gewinn / Verlust : 0.80

Neuer Kontostand : 12.40

Antworten Total: 19 (falsche: 4 / richtige: 15)

1 2 3 4 5 6 7

Gehe zur nächsten Frage, wenn Du bereit bist

Antwort (Erläuterung)

Deine Antwort ist richtig !

Ein Belemnit ist ein ausgestorbener Kopffüßler, der vor allem in den Meeren der Jura- und Kreidezeit vorkam. Ihre Gehäuse haben sich oft als Fossilien erhalten.

Weiter zu Frage Nr. 20

Abbildung 4.26: Schritt 7 (Lösung und Erläuterung ansehen und zur nächsten Aufgabe überge-

Frage

Wie heisst die bedeutendste hinduistische Tempelanlage auf Bali?

Prambanan	A
Pura Besakih	B
Mengwi	C
Borobudur	D

Abbildung 4.27: Frage und Antwort-Alternativen

Telefonjoker

Nr	Kosten	Erreichbarkeit	Zuverlässigkeit	Tipp
1	100 Pkte	18 %	69 %	<input type="checkbox"/>
2	100 Pkte	91 %	38 %	<input type="checkbox"/>
3	100 Pkte	13 %	51 %	<input type="checkbox"/>
4	100 Pkte	28 %	86 %	<input type="checkbox"/>
5	100 Pkte	64 %	61 %	<input type="checkbox"/>
6	100 Pkte	61 %	91 %	<input type="checkbox"/>

Bisherige Ausgaben : 0 Pkte

Mögliche Einnahme : 1'000 Pkte

Abbildung 4.28: Telefonjoker

Kontostand (Punkte) - Phase 1

Einnahme bei richtiger Antwort :	<input type="text" value="1'000"/>
Bisheriger Kontostand :	<input type="text" value="5'000"/>
Ausgaben bei aktueller Frage :	<input type="text" value="300"/>
Einnahme bei aktueller Frage :	<input type="text" value="1'000"/>
<hr/>	
Resultierender Gewinn / Verlust :	<input type="text" value="700"/>
Neuer Kontostand :	<input type="text" value="5'700"/>
<hr/>	
Antworten Total:	<input type="text" value="8"/> (falsche: <input type="text" value="2"/> / richtige: <input type="text" value="6"/>)

Abbildung 4.29: Kontostand und andere Angaben

Die Reihenfolge in der die Antwort-Alternativen präsentiert wurden, war wiederum völlig zufällig und wurde erst während des Versuchsdurchgangs festgelegt (Abbildung 4.27). Das gleiche galt für die Präsentations-Reihenfolge der einzelnen Joker innerhalb einer Joker-Gruppe. Auch diese Reihenfolge war völlig zufällig und wurde erst während des Versuchsdurchgangs festgelegt. Erst nachdem die Reihenfolge feststand, wurden die Joker von oben nach unten durchnummeriert (Abbildung 4.28). Dank der permanenten Einblendung der Kontoübersicht hatten die Versuchspersonen in beiden Phasen jederzeit den Überblick über eine Reihe wichtiger Daten wie z.B. ihre aktuellen Ausgaben, ihren momentanen Kontostand und die Anzahl der bisher richtig und falsch gelösten Aufgaben (Abbildung 4.29).

4.6 Vorversuch zur Auswahl der Wissensfragen

Um das Experiment durchführen zu können, musste ein entsprechender Aufgabenpoll selektiert schwieriger Fragen zusammengestellt werden. Dies geschah in einem eigenen Vorversuch.

4.6.1 Auswahl des Aufgaben-Basismaterials (Fragen und Antwort-Alternativen)

In einem ersten Schritt wurden ca. 600 schwierige Fragen aus dem Buch "Wissensportal Allgemeinwissen" (Edbauer, 2001) ausgewählt. Diese wurden zusammen mit den vier dort aufgeführten Antwort-Alternativen sowie dem ebenfalls aufgeführten kurzen Erläuterungstext manuell elektronisch erfasst. Die Erläuterungstexte wurden lediglich sinngemäss übernommen und auf eine maximale Länge von 255 Zeichen redigiert.

The screenshot displays a quiz interface with a light blue background. At the top, under the heading "Frage", is the question: "Welcher Krieg gilt bereits als 'Erster' Weltkrieg?". Below the question are four numbered options, each in a light blue box with a thin border:

- 1 Österreichischer Erbfolgekrieg
- 2 Siebenjähriger Krieg
- 3 Erster Nordischer Krieg
- 4 Spanischer Erbfolgekrieg

Below the options is a section titled "Antwort-Sicherheit" containing five buttons for confidence levels: "völlig geraten", "ziemlich unsicher", "eher sicher", "ziemlich sicher", and "absolut sicher". At the bottom of the interface, there is a "Kommentar" button on the left, a small box containing the number "1" in the center, and a "Nächste Frage >>" button on the right.

Abbildung 4.30: PreTest1 (Vorexperiment zum Testen der 600 Wissensfragen)

4.6.2 Vorversuch

Um eine Auswahl möglichst schwieriger Fragen zu erhalten, testeten wir die 600 Wissensfragen in einem Vorexperiment ("PreTest1", programmiert mit Visual Basic 6, Abbildung 4.30) an insgesamt 16 Versuchspersonen. Die Versuchspersonen durften den Vortest zu jedem beliebigen Zeitpunkt abbrechen. Die Reihenfolge der Fragen wurde vom Programm für jede Versuchsperson zufällig neu generiert. Die Versuchspersonen erhielten während des Vortests kein unmittelbares Feedback über die Korrektheit der von ihnen abgegebenen Antworten. Sie wurden jedoch am Ende des Tests darüber informiert, wie viele der Fragen sie richtig und wie viele sie falsch beantwortet hatten.

Damit für jede Frage mindesten 10 Ratings vorlagen, wurde der Fragepool bei den letzten beiden Versuchspersonen jeweils reduziert auf genau diejenigen Fragen, die bisher weniger als zehnmal beantwortet worden waren. Diese beiden letzten Personen mussten jeweils alle im reduzierten Fragepool verbliebenen Fragen beantworten.

Der Vorversuch hatte nebenbei den Zweck, allfällige Schreibfehler (sowie auch inhaltliche Fehler bzw. missverständliche Fragen, etc.) zu detektieren. Die Versuchspersonen wurden deshalb gebeten, Schreibfehler und inhaltliche Fehler zu melden. Dazu konnten sie im Bedarfsfall zu jeder Aufgabe direkt am Bildschirm einen Kommentar eingeben und die Fragen bzw. die einzelnen Antwort-Alternativen markieren, bei denen sie vermeintliche oder tatsächliche Fehler entdeckt hatten. Nachdem alle Versuchspersonen den Pretest absolviert hatten, erhielt jede per Email ein vollständiges Feedback zu den von ihr gelösten Aufgaben.

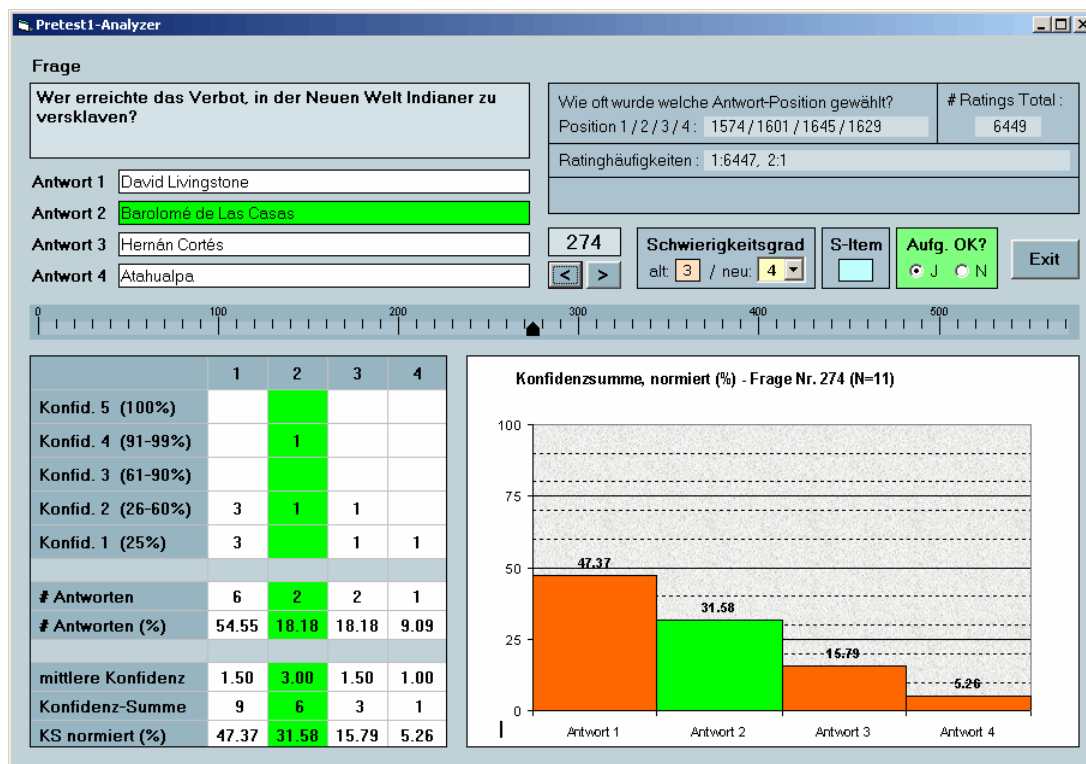


Abbildung 4.31: PreTest1-Analyzer (Programm für Analyse der Antwortverteilung in PreTest1)

4.6.3 Antwortmuster-Analyse

Zwecks endgültiger Auswahl der Wissensfragen wurde eine Antwortmuster-Analyse durchgeführt: Mittels eines Analyse-Programms zur Visualisierung und Bewertung der Antwortverteilung (Abbildung 4.31) liess sich entscheiden, welche der insgesamt 600 Fragen im Fragepool verbleiben sollen und welche nicht.

Die insgesamt vier Entscheidungskriterien waren wie folgt definiert: 1. Mindestens drei der vier Alternativen sind von zumindest einer Person gewählt worden. 2. Maximal 1 Versuchsperson hat die richtige Antwort sicher gewusst (Konfidenz = 5). 3. Maximal 50% der Versuchspersonen haben auf die richtige Antwort getippt. 4. Ausschluss aller Fragen (abgesehen von zwei Ausnahmen), bei denen zwei oder mehr Versuchsperson auf eine falsche Antwort bei gleichzeitig maximaler Antwortsicherheit (Konfidenz = 5) getippt haben.

4.6.4 Definitiver Fragepool

Am Ende dieses Prozesses verblieben insgesamt 140 schwierige Wissensfragen (inklusive je 4 Antwortalternativen und einem Erläuterungstext pro Frage) im Fragepool für das Hauptexperiment. In Tabelle 4.5 sind auszugsweise einige der 137 Wissensfragen (inklusive zugehöriger Antwortalternativen) zu sehen, mit denen die Versuchspersonen im Verlauf des Hauptdurchgangs des Experiments konfrontiert wurden. Die korrekte Antwort ist jeweils mit einem X in der letzten Spalte markiert. Jede Versuchsperson hatte 73 derartiger Wissensfragen zu beantworten. Im Probedurchgang mussten alle Versuchspersonen dieselben drei Fragen (Tabelle 4.6) beantworten, weil die Telefonjoker-Einsätze akustisch simuliert und dadurch fragenspezifisch vorbereitet werden mussten. Die Präsentationsreihenfolge der Fragen war ebenfalls für alle Versuchspersonen gleich (PD1, PD2, PD3). Anschliessend folgten 53 zufällig ausgewählte Fragen für Phase 1 und 20 zufällig ausgewählte Fragen für Phase 2 des Hauptdurchgangs. Pro Versuchsperson wurde jede der insgesamt 76 Fragen nur ein einziges Mal gestellt (= Ziehung ohne Zurücklegen). Tabelle 4.7 enthält eine systematische Auflistung der Parameter (Grösse des Fragepools, Reihenfolge der Aufgaben, etc.) für jeden Durchgang bzw. für jede Phase.

Tabelle 4.5: Auszug aus dem Fragepool für den Hauptdurchgang des Experiments

5	Wo entstand das Königreich der Marathen?	1	Indien	X
		2	Malaysia	
		3	Siam	
		4	Persien	
18	Wer stellte 1898 erstmals Heroin her?	1	Hoechst	
		2	Bayer	X
		3	Robert Koch	
		4	John Franklin-Enders	
20	Welches deutsche Markenprodukt wurde 1893 auf der Weltausstellung in Chicago vorgestellt?	1	Sauerkraut	
		2	Weisswurst	
		3	Persil	
		4	Leibniz-Keks	X
27	Welcher altägyptische Gott gilt als Schöpfer der Hieroglyphen?	1	Seth	
		2	Thot	X
		3	Anubis	
		4	Indra	
29	Was ist das "Debora-Lied"?	1	Sapphische Ode	
		2	Talmud-Dichtung	
		3	Bibel-Verse	X
		4	Kabbala-Sprüche	
30	Wen verehrten die alten Perser als den "Weisen Gott"?	1	Adad	
		2	Ischum	
		3	Ahuramazda	X
		4	Nergal	
31	Wer entwickelte die Impfstoffe gegen die Infektionskrankheiten Diphtherie und Tetanus?	1	Fridtjof Nansen	
		2	Emil von Behring	X
		3	Edward Jenner	
		4	Louis Pasteur	
36	Wie heisst die bedeutendste hinduistische Tempelanlage auf Bali?	1	Mengwi	
		2	Pura Besakih	X
		3	Borobudur	
		4	Prambanan	
42	Wo erschien 1605 die erste Zeitung?	1	Leipzig	
		2	London	
		3	Neu Amsterdam	
		4	Strassburg	X
45	Welche schwarze Mezzosopranistin trat 1961 als erste farbige Sängerin in Bayreuth auf?	1	Leontyne Price	
		2	Barbara Hendricks	
		3	Grace Bumbry	X
		4	Jessey Norman	

Tabelle 4.6: Die drei Wissensfragen des Probedurchgangs (inkl. zugehöriger Antwortalternativen)

PD1	Wer erfand 1800 die erste Batterie?	1	James Watt	
		2	Alessandro Volta	X
		3	André Marie Ampère	
		4	William Henry	
PD2	Wie hiess Tasmanien bis 1853?	1	Melanesien	
		2	Chatham-Inseln	
		3	Van-Diemens-Land	X
		4	Georgina	
PD3	Welcher jüdisch-spanische Schriftsteller schrieb in deutscher Sprache den Roman "Die Blendung" (1935)?	1	Elias Canetti	X
		2	Ortega y Gasset	
		3	Camilo José Cela	
		4	Franz Werfel	

Tabelle 4.7: Übersicht über die Phasen-Parameter

	<i>Probendurchgang</i>	<i>Hauptdurchgang, Phase 1</i>	<i>Hauptdurchgang, Phase 2</i>
<i>Aufgaben-Basismenge (Pool)</i>	3	137 (= 140 - 3)	84 (= 140 - 3 - 53)
<i>Aufgabenmenge pro Vp</i>	3	53 (45 + 5 + 3)	20 (2 x 10)
<i>Aufgaben-Reihenfolge</i>	fest vorgegeben	zufällige Reihenfolge	zufällige Reihenfolge
<i>Anordnung der Antwortalternativen</i>	fest vorgegeben	zufällige Anordnung	zufällige Anordnung
<i>Cue-Sets: Basismenge (Pool)</i>	3	53	50 (15 pro Strategie, überlagernd)
<i>Cue-Sets: Benötigte Anzahl pro Vp</i>	3	53	20 (2 x 10)
<i>Cue-Sets: Selektion</i>	alle 3	alle 53	2 x 10 (abh. von Strategie- Präferenz)
<i>Cue-Sets: Reihenfolge</i>	fest vorgegeben: Nr 1, Nr 2, Nr 3	ausbalanciert über alle Vpn (vordefiniert für jede Vp)	ausbalanciert über alle Vpn (vordefiniert für jede Vp)
<i>Cue-Anordnung innerhalb Cue-Sets</i>	fest vorgegeben	zufällige Anordnung	zufällige Anordnung
<i>Erzeugung der Cue-Information (Joker-Tipps)</i>	Tipps sind vor- definiert (wegen zugehörigen Audio-Samples)	Tipps werden on-the-fly erzeugt (gemäss Validität & Diskriminationsrate)	Tipps werden on-the-fly erzeugt (gemäss Validität & Diskriminationsrate)
<i>Anfangsbonus</i>	1000 Punkte	1'000 Punkte	CHF 2.00
<i>Cueskosten</i>	100 Punkte	100 Punkte	CHF 0.20
<i>Mögliche Gewinnsumme</i>	1'000 Punkte	1'000 Punkte	CHF 2.00

4.7 Erzeugung der Cues und Konstruktion der Cue-Sets

Durch den Vergleich der individuellen Informationssuch-Reihenfolge mit den strategiekonformen Informationssuch-Reihenfolgen der fünf untersuchten Suchstrategien (Validität, Diskriminationsrate, Success, Usefulness sowie V+D als rechnerisch vereinfachte Kombination) soll für jede Aufgabe bestimmbar sein, ob eine Versuchsperson darin eine der fünf untersuchten Suchstrategien angewendet hat und wenn ja welche. Indem anschliessend die Häufigkeiten der einzelnen detektierten Suchstrategien über alle von einer Versuchsperson gelösten Aufgaben ausgezählt werden, stellt das Programm fest, welche der untersuchten Strategien von dieser Person präferiert wird.

Für die insgesamt 73 Aufgaben der beiden Hauptteile des Experiments werden entsprechend viele Telefonjoker-Gruppen (Cue-Sets) à je 6 Joker (Cues) benötigt. Die Cues müssen in ihren Validitäten und Diskriminationsraten jeweils so zu einem Cue-Set zusammengestellt werden, dass spätestens mit dem Kauf des zweiten Cues eindeutig bestimmt werden kann, ob die Versuchsperson eine der fünf getesteten Suchstrategien angewendet hat und falls ja, welche.

Das einmal im Programm zusammengestellte Cue-Set ist fix, d.h. jede Versuchsperson erhält im gesamten Verlauf des Experiments die gleichen 73 Cue-Sets angeboten, allerdings in randomisierter Reihenfolge. Die Anwendung einer bestimmten Suchstrategie wird nur dann detektiert, wenn eine Versuchsperson die entsprechenden Cues in der exakt strategiekonformen Reihenfolge aufdeckt. Damit ist ein sehr strenges Kriterium an die Detektion der Suchstrategien angelegt.

Jedes Cue-Set ist so konstruiert, dass darin bezüglich aller fünf untersuchten Suchstrategien keine zwei gleich guten Cues auftauchen. Dadurch ist für jede Suchstrategie eine eindeutige Konsultationsreihenfolge definiert. Es lässt sich jedoch bei der Konstruktion der Cue-Sets nicht vermeiden, dass einige der strategiekonformen Reihenfolgen innerhalb eines Cue-Sets partiell oder sogar vollständig kongruent sind. Die Detektion muss jedoch auch funktionieren, wenn die Versuchspersonen in erster Linie One-Reason Decision Making betreiben. In diesen Fällen stehen zur Festlegung einer Suchstrategie i.d.R. nur relativ wenige konsultierte Cues zur Verfügung, so dass wir zur Erzeugung der Cue-Sets die Forderung einzuführen hatten, dass die eindeutige Detektion der Suchstrategien spätestens mit dem zweiten strategiekonformen Cue-Kauf möglich sein muss (Restriktion 1, siehe unten).

Mit einer einzigen Aufgabe lassen sich aber nur maximal drei verschiedene Suchstrategien innerhalb der ersten beiden Cue-Käufe voneinander unterscheiden (in den meisten Fällen sogar weniger). Deswegen gelten für das Cue-Set als Ganzes die folgenden Anforderungen: 1. Jede Suchstrategie muss genügend oft eindeutig detektiert werden können. 2. Jede Suchstrategie muss gleich oft eindeutig detektiert werden können.

Als Folge der Umsetzung dieser beiden Anforderungen kommt es notwendigerweise zu Schwerpunktsetzungen in der Gleichverteilung der Validitäten und Diskriminationsraten. (So entstehen bspw. sehr viel mehr Cues mit sehr hohen Validitäten als mit niedrigen oder mittelhohen). Für die Versuchspersonen sollte davon aber nichts zu spüren sein. Deshalb ist im Programm für eine Gleichverteilung der Validitäten und Diskriminationsraten bezüglich gleichgrosser, sich nicht überlappender Intervalle gesorgt, was in etwa einer Zufallsverteilung entspricht.

Ebenfalls aufgrund der beiden Anforderungen ist bereits a priori festzulegen, welche Cues eines Cue-Sets diskriminieren und welche nicht. Auch hier ist dafür gesorgt, dass die Anzahl diskriminierender Cues pro Cue-Set in etwa der Zufallsverteilung folgt – allerdings unter Berücksichtigung von Restriktion 6 (siehe weiter unten).

Die Cue-Werte (Joker-Tipps) werden hingegen erst während des Versuchsablaufs, unter Berücksichtigung der Zuverlässigkeit (Validität) und der Erreichbarkeit (Diskriminationsrate) des Jokers, mittels einer verdeckten Zufallsziehung vom Programm bestimmt. Vorausgesetzt ein Joker

kann telefonisch erreicht werden, erzeugt das Programm eine Zufallszahl zwischen 1 und 100. Ist diese Zufallszahl grösser als die Validität des Jokers (in Prozent), so tippt der Joker auf eine zufällig ausgewählte falsche Antwort. Bei einer Zufallszahl kleiner gleich seiner Validität tippte er auf die korrekte Antwort.

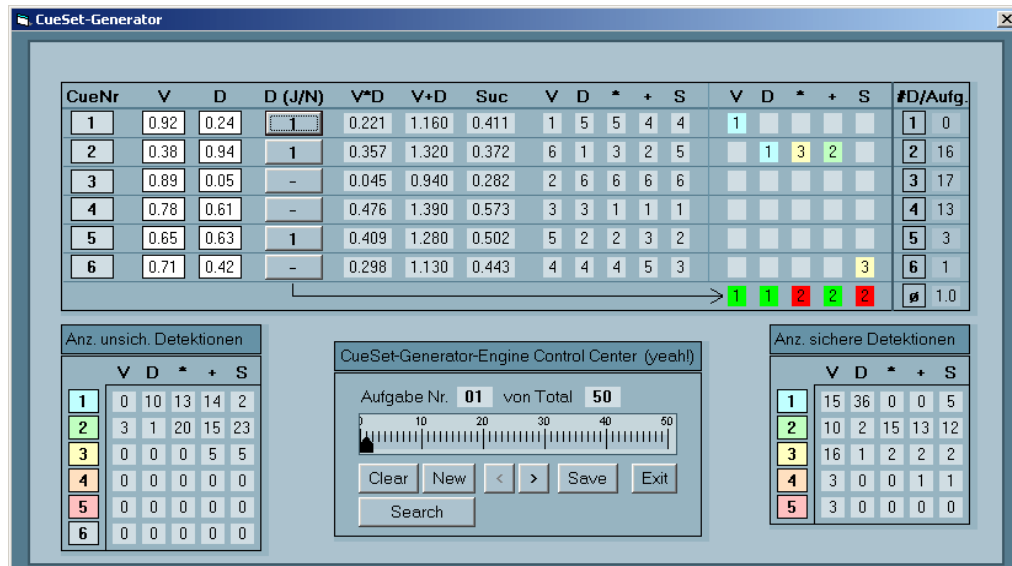


Abbildung 4.32: CueSet-Generator (Programm zur Erzeugung von Cue-Sets)

Die halbautomatische Erzeugung der Cue-Sets mittels eines speziell dafür konstruierten Generator-Programms ('CueSet-Generator', siehe Abbildung 4.32) basiert auf der zufallsgesteuerten Generierung von sechs Cues und unterliegt den folgenden sechs Restriktionen:

Restriktion 1: Die mit einem bestimmten Cue-Set detektierbaren Suchstrategien werden spätestens mit der zweiten strategiekonformen Cue-Wahl eindeutig detektiert (d.h. sie können spätestens mit der zweiten strategiekonformen Cue-Wahl von allen anderen Strategien unterschieden werden).

Restriktion 2: Alle Cues innerhalb eines Cue-Sets unterscheiden sich, paarweise miteinander verglichen, bezüglich ihrer Güte um einen minimalen Betrag (Delta, d) voneinander. Für die V- und die D-Strategie beträgt $d = 0.01$ und für alle anderen Strategien (V+D, V*D und SUC) ist $d = 0.03$.

Restriktion 3: Jeder Cue, d.h. jede beliebige Kombination von V und D, kommt in allen erzeugten Cue-Sets zusammen maximal einmal vor.

Damit ist sichergestellt, dass die Cues voneinander unabhängig sind und dass kein Lernen stattfinden kann.

Restriktion 4: Die höchste Zuverlässigkeit (Validität) eines Jokers in einem Cue-Set beträgt mindestens 0.88. Diese Restriktion gilt nur für die drei Cue-Sets im Probedurchlauf und für alle

Standard-Cue-Sets in Phase 1 und in Phase 2 (vgl. Tabelle 4.8). Die Tabelle zeigt die als Resultat der sechs Restriktionen eingestellten Minima der maximalen Validitäten (minimales V_{\max}) der Cue-Sets, abhängig vom Cue-Set-Typ.

Tabelle 4.8: Geforderte Minima der maximalen Cue-Set-Validitäten

Cue-Set-Typ	Phase	Anzahl Sets	minimales V_{\max}
Probedurchgang-Set	-	3	0.90
Standard-Set	1	45	0.88
Füll-Set	1	5	0.82
Validitäts-Test-Set	1	3	0.26 ¹
Standard-Set	2	20	0.88

¹ Nur für genau einen der Cues darf $V \geq 0.26$ sein.

Restriktion 5: Die Wertebereiche für die zufällige Erzeugung der Erreichbarkeit (D) und der Zuverlässigkeit (V) eines Cues sind restringiert, da keine Extremwerte und insbesondere keine zu kleinen Validitäten im Experiment vorkommen sollen.

restringierter Bereich für V: 0.35 – 0.95

restringierter Bereich für D: 0.05 – 0.95

Diese Restriktion gilt natürlich nicht für die drei Validitäts-Test-Sets, mit denen überprüft wird, ob eine Versuchsperson das Konzept der Validität verstanden hat.

Restriktion 6: In jedem Cue-Set diskriminieren mindestens zwei der sechs Cues.

Anmerkung zu Restriktion 2. Grössere Deltas führen zu einer potentiell besseren Differenzierung der Cue-Güte innerhalb eines Cue-Sets. Damit erhöhen wir die Wahrscheinlichkeit, dass unter ähnlich guten Cues der bezüglich der verwendeten Suchstrategie objektiv bessere Cue als solcher erkannt und gewählt wird. Gleichzeitig sinkt dadurch die Wahrscheinlichkeit, dass Cues innerhalb der individuellen Suchreihenfolge aufgrund von Berechnungsfehlern miteinander vertauscht werden und damit eine saubere Strategiedetektion verunmöglichen. Leider können wir keine beliebig grossen Deltas fordern: Umso grösser das Delta einer Strategie in einem Cue-Set ist, desto weniger Strategien können wir mit dem betreffenden Cue-Set voneinander unterscheiden bzw. eindeutig detektieren.

Anmerkung zu Restriktion 4. Für die fünf Füll-Sets zur globalen Ausbalancierung der Validitäten und Diskriminationsraten ist diese Restriktion nicht beachtet. Ebenfalls ausser Kraft tritt diese Restriktion bei der Konstruktion der drei Cue-Sets für den Validitäts-Test: Bei diesen drei Cue-Sets weist jeweils genau einer der sechs Cues eine Validität grösser 0.25 auf. Die Validitäten der verbleibenden fünf Cues sind alle kleiner gleich 0.25 und bewegen sich im Bereich von 0.15 bis 0.25.

4.8 Mittlere Diskriminationsrate der Cues

Die Telefonjoker-Sets (Cue-Sets) sind so zusammengestellt, dass pro Aufgabe jeweils mindestens 2 Cues (Telefonjoker) diskriminieren. Mit dieser Voreinstellung wird vermieden, dass die Versuchspersonen durch zu häufige Nicht-Erreichbarkeit von Jokern frustriert werden. Nach Berücksichtigung aller oben eingeführten Restriktionen zur sauberen Detektion der Suchstrategien resultiert über alle Cue-Sets eine gemittelte Diskriminationsrate von 0.52 (es diskriminieren also im Mittel 3.125 Cues pro Aufgabe).

Der verwendete Algorithmus zur Erzeugung von Cue-Sets mit mindestens zwei diskriminierenden Cues ist folgender:

1. Generiere ein Cue-Set mit einer zufälligen Anzahl an diskriminierenden Cues.
2. Behalte das Cue-Set, genau dann wenn mindestens zwei Cues diskriminieren und fahre fort bei Schritt 1, falls noch weitere Cue-Sets benötigt werden.
3. Verwirf das Cue-Set genau dann wenn weniger als zwei Cues diskriminieren.
4. Generiere so lange ein neues Cue-Set, bis ein Cue-Set mit genau (!) zwei diskriminierenden Cues resultiert.
5. Verwende das neue Cue-Set (mit genau zwei diskriminierenden Cues) anstelle des verworfenen Cue-Sets (mit weniger als zwei diskriminierenden Cues) und fahre fort bei Schritt 1, falls noch weitere Cue-Sets benötigt werden.

Die Verteilung der Anzahl diskriminierender Cues ist so aufgebaut, dass sie perfekt die Zufallsverteilung um eine Diskriminationsrate von 0.50 abbildet – freilich mit der oben eingeführten Beschränkung, dass in jeder Aufgabe mindestens zwei Joker erreichbar sein müssen. Im Folgenden wird gezeigt, wie die durchschnittliche Diskriminationsrate von 0.52 durch bestmögliche Annäherung an eine Zufallsverteilung um 0.50 zustande kommt: a_k sei die mittlere Wahrscheinlichkeit dafür, dass in einem Cue-Set k diskriminierende Cues vorhanden sind (mit anderen Worten: a_k ist die 'Auftretenswahrscheinlichkeit' für k diskriminierende Cues pro Cue-Set). a_k kann mit der untenstehenden Formel berechnet werden. Dabei gilt: $n=6$ (da ein Cue-Set insgesamt 6 Cues umfasst) und $p=0.5$ (da genau zwei Ereignisse möglich sind: Entweder der Cue diskriminiert oder er diskriminiert nicht):

Tabelle 4.9 zeigt die sich daraus ergebenden mittleren Auftretenswahrscheinlichkeiten a_k für jeweils k diskriminierende Cues pro Cue-Set bei zufallsverteilter Cue-Diskrimination (d.h. bei zufälliger Verteilung der Fälle 'Cue diskriminiert' bzw. 'Cue diskriminiert nicht' über alle Cues des Cue-Sets):

Tabelle 9: Auftretenswahrscheinlichkeiten für k diskriminierende Cues pro Cue-Set bei zufallsverteilter Cue-Diskrimination

k	Auftretenswahrscheinlichkeit (a_k)
6	1.5625 %
5	9.3750 %
4	23.4375 %
3	31.2500 %
2	23.4375 %
1	9.3750 %
0	1.5625 %

Tabelle 10: Auftretenswahrscheinlichkeiten für k diskriminierende Cues pro Cue-Set bei restringierter Verteilung von D

k	Auftretenswahrscheinlichkeit (a_k)
6	1.5625 %
5	9.3750 %
4	23.4375 %
3	31.2500 %
2	23.4375 %
1	9.3750 %
0	1.5625 %

Es ist leicht zu sehen, dass a_k symmetrisch um den Mittelwert verteilt ist und dass dieser bei $k=3$ liegt. Mit anderen Worten: In einem Cue-Set mit zufallsverteilter Cue-Diskrimination diskriminieren im Mittel genau drei Cues.

Bei Anwendung des weiter oben beschriebenen Algorithmus' zur Generierung von Cue-Sets mit mindestens zwei diskriminierenden Cues werden die beiden Fälle $k=0$ und $k=1$ nicht benötigt. Stattdessen werden nun mit den entsprechenden Wahrscheinlichkeiten (d.h. mit 1.5625 % resp. 9.3750 %) Cue-Sets mit jeweils genau zwei diskriminierenden Cues erzeugt.

Die neue Auftretenswahrscheinlichkeit für $k=2$ ergibt sich folglich, indem man die Auftretenswahrscheinlichkeiten der drei Fälle $k=0$, $k=1$ und $k=2$ aufsummiert. Damit erhält man für zwei diskriminierende Cues eine Auftretenswahrscheinlichkeit von 34.375 % (Tabelle 4.10).

Die Anzahl d der im Mittel diskriminierenden Cues pro Cue-Set beträgt damit neu 3.125. Diesen Wert erhält man, wenn man die jeweilige Anzahl k der diskriminierenden Cues pro Cue-Set mit ihrer Auftretenswahrscheinlichkeit a_k multipliziert und alle Produkte aufsummiert. Hier die Formel:

$$\begin{aligned}
 d &= \sum_{k=0}^6 (k \cdot a_k) \\
 &= 9.375 \% + 46.875 \% + 93.75 \% \\
 &\quad + 93.75 \% + 68.75 \% + 0 \% + 0 \% \\
 &= 312.5 \% = 3.125
 \end{aligned}$$

Da die Beziehung zwischen k und mittlerer Diskriminationsrate linear ist, ergibt sich für die im Mittel 3.125 diskriminierenden Cues pro Cue-Set eine mittlere Diskriminationsrate von 52.083 Prozent über alle Cues.

Eine alternative Möglichkeit, die mittlere Diskriminationsrate über alle Cues zu berechnen, besteht darin, die mittlere Diskriminationsrate pro k diskriminierende Cues zu ermitteln, mit der jeweiligen Auftretenswahrscheinlichkeit a_k zu gewichten und die gewichteten Diskriminationsraten aufzusummieren. Die a_k sind bereits bei der ersten Berechnungsvariante ermittelt worden und können Tabelle 4.9 bzw. Tabelle 4.10 entnommen werden. Tabelle 4.11 verdeutlicht die Situation bei zufallsverteilter, Tabelle 4.12 bei restringierter Cue-Diskrimination:

Tabelle 4.11: Mittlere Diskriminationsrate (DR) über alle Cues bei zufallsverteilter Cue-Diskrimination

k	mittlere DR	a_k (Gewichtung)	gewichtete DR
6	6/6 = 100.00 %	1.56 %	1.56 %
5	5/6 = 83.33 %	9.38 %	7.81 %
4	4/6 = 66.66 %	23.44 %	15.63 %
3	3/6 = 50.00 %	31.25 %	15.63 %
2	2/6 = 33.33 %	23.44 %	7.81 %
1	1/6 = 16.66 %	9.38 %	1.56 %
0	0/6 = 0.00 %	1.56 %	0.00 %
Summe: Mittlere DR über alle Cues :			50.00 %

DR = Diskriminationsrate

Tabelle 4.12: Mittlere Diskriminationsrate (DR) über alle Cues bei restringierter Cue-Diskrimination

k	mittlere DR	a_k (Gewichtung)	gewichtete DR
6	6/6 = 100.00 %	1.56 %	1.56 %
5	5/6 = 83.33 %	9.38 %	7.81 %
4	4/6 = 66.66 %	23.44 %	15.63 %
3	3/6 = 50.00 %	31.25 %	15.63 %
2	2/6 = 33.33 %	34.38 %	11.46 %
1	1/6 = 16.66 %	0.00 %	0.00 %
0	0/6 = 0.00 %	0.00 %	0.00 %
Summe: Mittlere DR über alle Cues :			52.08 %

DR = Diskriminationsrate

4.9 Bestimmung der in Phase 2 zu testenden Suchstrategien

Am Ende von Phase 1 des Hauptdurchgangs musste das Programm ermitteln, welche zwei Suchstrategien die Versuchsperson in dieser Phase am stärksten präferierte, um sie in Phase 2 des Hauptdurchgangs noch einmal gegeneinander testen zu können. Primäres Kriterium zur Bestimmung dieser beiden Strategien war ihre Anwendungshäufigkeit (Schritte 1 und 2 in Tabelle 4.13).

Tabelle 4.13: Prozedur-Schritte zur Bestimmung der Strategie-Präferenz-Rangfolge

Schritt-Nr.	Beschreibung
1	Bestimmung der Anwendungshäufigkeit (AH) jeder Strategie
2	daraus Strategie-Rangfolge RAH bilden basierend auf den Anwendungshäufigkeiten
3	Tiebreaker1 (TB1): Zählerstände bestimmen gemäss TB1-Kriterium
4	daraus Strategie-Rangfolge RT1 bilden basierend auf den TB1-Zählerständen
5	Tiebreaker2 (TB2): Zählerstände bestimmen gemäss TB2-Kriterium
6	daraus Strategie-Rangfolge RT2 bilden basierend auf den TB2-Zählerständen
7	Präferenz jeder Strategie berechnen: $Präf_i = (100 * RAH_i) + (10 * RT1_i) + RT2_i$
8	daraus Strategie-Rangfolge bilden, basierend auf $Präf_i$
9	Die beiden am stärksten präferierten Strategien auswählen (Rang 1 und Rang 2)

Tabelle 4.14: Auflösung eines Gleichstands dank Tiebreaker 1

Aufgaben-Nummer	VAL	DIS	V*D	V+D	SUC	Zeilensumme
1	1	0	0	0	0	1
2	1	0	0	0	0	1
3	0	0	1	0	1	2
4	0	1	0	0	1	2
5	1	0	0	0	0	1
6	0	1	0	0	1	2
7	1	0	0	0	1	2
Anw.-Häufigkeit (AH)	4	2	1	0	4	
Tiebreaker1 (TB1)	3	0	0	0	0	
Tiebreaker2 (TB2)	5	4	2	0	8	
Umgekehrter Rang AH	4	3	2	1	4	Gewichtung: 100
Umgekehrter Rang TB1	2	1	1	1	1	Gewichtung: 10
Umgekehrter Rang TB2	2	3	4	5	1	Gewichtung: 1
Präferenz (Präf)	422	313	214	115	411	
Präferenzrang	1	3	4	5	2	→ VAL gewinnt!

Dieses Kriterium allein reichte aber nicht in allen denkbaren Fällen aus, denn nach der Auszählung der Häufigkeiten konnte es ja durchaus einmal zu einem Gleichstand zwischen zwei oder mehreren Strategien kommen. Weil das Design aber vorsah, immer genau zwei Strategien in Phase 2 gegeneinander zu testen, führten wir zwei sukzessiv anzuwendende Tiebreaker ein (gleich zwei, um möglichst sicher zu gehen, dass ein allfälliger Gleichstand tatsächlich aufgelöst wird; Schritte 3 bis 6 in Tabelle 4.13). Hätte ein allfälliger Gleichstand auch unter Berücksichtigung beider Tie-Break-Kriterien nicht aufgelöst werden können, so wäre die Rangordnung der Präferenz innerhalb der gleichständigen Strategien durch eine Zufallsauswahl festgelegt worden. Für die Tiebreaker werden alle Aufgaben, die eine Person gelöst hatte, noch einmal einzeln betrachtet:

Kriterium von Tiebreaker1: Der Zähler für eine bestimmte Strategie wird genau dann um 1 erhöht, wenn diese Strategie in der betrachteten Aufgabe die einzige ist, die überhaupt in Frage kommt, d.h. wenn die Zeilensumme der Strategie-Indikatoren gleich 1 ist (Tabelle 4.14). Je grösser der Zählerstand nach Betrachtung aller Aufgaben war, desto grösser die Bedeutung dieser Strategie.

Kriterium von Tiebreaker2: Der Zähler für eine Strategie wird genau dann um einen bestimmten Betrag erhöht, wenn die Strategie in der betrachteten Aufgabe in Frage kommt (d.h. in der Datenbank mit einer "1" markiert ist). Der Betrag, um den der Zähler erhöht wird, ist gleich der Anzahl der in dieser Aufgabe insgesamt in Frage kommenden Strategien. Mit anderen Worten: Der Zähler wird um den Betrag der Zeilensumme erhöht (Tabelle 4.15). Je kleiner der Zählerstand nach Betrachtung aller Aufgaben war, desto grösser die Bedeutung dieser Strategie. Die Präferenz einer Strategie konnte nun gemäss der folgenden Formel berechnet werden:

$$\text{Präf}_i = (100 * \text{RAH}_i) + (10 * \text{RT1}_i) + \text{RT2}_i$$

Dabei ist RAH_i der Rang von Strategie i bezüglich dem primären Präferenz-Kriterium, d.h. der Anwendungshäufigkeit. $RT1_i$ und $RT2_i$ sind die Ränge von Strategie i bezüglich der beiden Tie-Break-Kriterien. Auf diese Weise konnten die beiden wichtigsten angewendeten Strategien in Phase 1 bestimmt und in Phase 2 gegeneinander getestet werden.

Tabelle 4.15: Auflösung eines Gleichstands dank Tiebreaker 2

Aufgaben-Nummer	VAL	DIS	V*D	V+D	SUC	Zeilensumme
1	0	0	1	0	0	1
2	0	0	0	0	1	1
3	1	0	0	0	1	2
4	0	0	1	0	1	2
5	1	0	1	1	0	3
Anw.-Häufigkeit (AH)	2	0	3	1	3	
Tiebreaker1 (TB1)	0	0	1	0	1	
Tiebreaker2 (TB2)	5	0	6	3	5	
Umgekehrter Rang AH	3	1	4	2	4	Gewichtung: 100
Umgekehrter Rang TB1	1	1	2	1	2	Gewichtung: 10
Umgekehrter Rang TB2	2	4	1	3	2	Gewichtung: 1
Präferenzwert (Präf)	312	114	421	213	422	
Präferenzrang	3	5	2	4	1	→ SUC gewinnt!

4.10 Performance-Abschätzung

Zur Erhöhung der ökologischen Validität des Experiments werden die Versuchspersonen leistungskontingent bezahlt – analog vergleichbarer Experimente in den Wirtschaftswissenschaften, wo die leistungskontingente Bezahlung der Versuchspersonen zu den üblichen Experimental-Standards gehört. Eine adäquate Möglichkeit zur Umsetzung leistungskontingenter Bezahlung besteht darin, die Versuchspersonen in Phase 1 des Hauptdurchgangs um einen Punktbetrag und in Phase 2 um einen realen CHF-Betrag spielen zu lassen. Durch die simple Multiplikation der in den beiden Phasen erspielten Beträge, erhalten wir eine einfache, auch für die Versuchspersonen leicht nachvollziehbare Möglichkeit zur Berechnung des Spielgewinns. Zudem sensibilisieren wir die Versuchspersonen damit in Phase 2 nochmals auf die Bedeutung ihrer Entscheidungen: Der hemmungslose Einsatz vieler Joker und die falsche Beantwortung einer Frage resultieren in höheren Kosten resp. in einem ausbleibenden Gewinn. Indem eine Versuchsperson in Phase 2 um einen realen CHF-Betrag spielt, erlebt sie einen entgangenen oder geschmäleren Gewinn realer, d.h. intensiver und schmerzhafter, als wenn sie bloss um Punkte gespielt hätte. Dementsprechend sollte sich eine Versuchsperson in Phase 2 nochmals stärker auf die von ihr verfolgte Suchstrategie und die damit verbundene spezifische Bewertung der Cues konzentrieren.

Zur Berechnung der Performance, also des am Ende des Experiments bar ausbezahlten Spielgewinns, werden die in den beiden Phasen erspielten Beträge einfach miteinander multipliziert. Für den in Phase 1 erspielten Punktbetrag ist allerdings noch ein geeigneter Divisor nötig. Die Performance (P) einer Versuchsperson berechnet sich schliesslich wie folgt:

$$\begin{aligned}
 P &= (\text{Punktestand in Phase 1} / \text{Divisor}) * \text{CHF-Gewinnbetrag in Phase 2} \\
 &= F_1 * \text{CHF-Gewinnbetrag in Phase 2}
 \end{aligned}$$

Ausserdem ist die Kosten/Gewinn-Ration festzulegen, d.h. das Verhältnis zwischen dem Einsatz eines Telefonjokers und dem Gewinn bei richtiger Antwort. Gemäss früherer Experimente (Läuge, Hausmann & Christen, 2005) hat sich ein Kosten-Gewinn-Verhältnis von etwa 1:10 als favorabel erwiesen.

Tabelle 4.16: Joker-Kosten und Gewinn-Berechnungen für die beiden Phasen des Experiments

	Phase 1 {Punkte}	Phase 2 {CHF}
Anfangsbonus (Kontostand zu Beginn der Phase)	1'000	2
Kosten pro Jokerkauf	100	0.1
Maximal mögliche Ausgaben pro Aufgabe (= Einsatz von 6 Jokern)	600	0.6
Möglicher Gewinnbetrag pro Aufgabe (Einnahme für richtige Antwort)	1'000	1
Maximal erreichbarer Punktestand bzw. CHF-Betrag (inkl. Anfangsbonus)	54'000	22
Durch Raten im Mittel erreichbarer Punktestand/CHF-Betrag (inkl. Anfangsbonus)	14'250	7

Tabelle 4.16 enthält eine Übersicht über die definitiv festgelegten Punkte-Beträge für Phase 1 sowie die CHF-Beträge für Phase 2. Ausserdem informiert sie über die maximalen sowie die durch Raten erspielbaren Punkte-/CHF-Beträge innerhalb einer Phase. Der im Mittel durch Raten erreichbare Betrag liegt bei genau einem Viertel des ohne Anfangsbonus erreichbaren maximalen Betrags (53'000 Punkte bzw. CHF 20.-), wobei der Bonus nach der Division wieder addiert werden muss: $53'000/4 + 1000 = 14'250$ bzw. $20.-/4 + 2.- = 7.-$.

Zur Performance-Berechnung ist noch ein geeigneter Divisor festzulegen. Die Anwendung des Divisors auf den in Phase 1 erspielten Punktebetrag sollte zu einem Faktor F_1 führen, der im für die Versuchsperson besten Fall deutlich grösser als 1 ist. Dieser für die Versuchsperson beste Fall wäre der, dass sie alle Fragen richtig beantwortet, ohne einen einzigen Joker einsetzen zu müssen. Wir können allerdings mit sehr grosser Wahrscheinlichkeit ausschliessen, dass dieser Fall auch nur annähernd eintreten wird. Indem wir dafür sorgen, dass F_1 im besten Fall deutlich grösser als 1 (aber nicht grösser als 3) wird, erreichen wir, dass die Ausprägung von F_1 im mittleren Fall immer noch einen für die Versuchspersonen akzeptablen Wert (d.h. einen Wert zwischen 0.5 und 1.5) annimmt. Der Divisor muss folglich kleiner als 54'000 und grösser oder gleich 18'000 sein. Dadurch ist sichergestellt, dass F_1 im besten Fall grösser als 1 und kleiner gleich 3 ist. Das Programm verwendet schliesslich ein F_1 von 3 und damit eine Divisor-Ausprägung von 18'000 ($54'000 / 18'000 = 3$).

Tabelle 4.17: Wichtige Annahmen für die Performance-Berechnung (basierend auf bisherigen Daten)

	Annahme für Phase 1	Annahme für Phase 2
Anzahl im Mittel eingesetzter Telefonjoker	2	2
Anteil der im Mittel richtig gelösten Aufgaben pro Vp (x_i)	59 %	66 %
Im Mittel anfallende Kosten für die eingesetzten Joker	220 Punkte	CHF 0.23
Mittlerer Gewinn für eine richtig gelöste Aufgabe (y_i)	780 Punkte	CHF 0.77
Im Mittel erspielter Gewinn pro Aufgabe ($x_i y_i - u_i v_i$)	375 Punkte	CHF 0.43

Tabelle 4.18: Performance-Berechnung (basierend auf den getroffenen Annahmen)

	Berechnung	exakter Betrag	gerundeter Betrag
Maximal mögliche Performance einer Versuchsperson	$(54'000 / 18'000) * 22.-$	CHF 66.00	CHF 66.-
Mittlere Performance einer Versuchsperson bei reinem Raten	$(14'250 / 18'000) * 7.-$	CHF 5.54	CHF 6.-
Mittlere Performance einer Versuchsperson (ohne Raten)	$(1000 + (53 * 375)) / 18'000 * (2 + (20 * 0.43))$	CHF 12.29	CHF 13.-
Totale Kosten total (bei mittlerer Performance ohne Raten)	$24 * 12.29$ bzw. $24 * 13.-$	CHF 294.96	CHF 312.-

Zu einer realistischen (das normale Budget für derartige Experimente nicht sprengenden noch die Versuchspersonen zu kurz kommen lassenden) Bezahlung sind für die Berechnung der mittleren Performance einer Versuchsperson noch einige zusätzliche Annahmen zu machen (siehe Tabelle 4.17): Es ist von vornherein abzuschätzen, wie viele Joker eine Person im Mittel einsetzen wird und wie viele Fragen sie korrekt beantwortet. Wir nehmen an, dass im Mittel zwei Telefonjoker eingesetzt werden. Unsere bisherigen Daten ergeben, dass die Versuchspersonen 59% der Aufgaben von Phase 1 und 66% der Aufgaben aus Phase 2 korrekt beantworten. Tabelle 4.18 enthält eine Übersicht über die mittlere und die maximal mögliche Performance einer Versuchsperson.

Die mittlere Performance (P) einer Versuchsperson kann gemäss der folgenden Formel berechnet werden:

$$P = \frac{(B_1 + N_1(x_1 y_1 - u_1 v_1))}{D} \cdot (B_2 + N_2(x_2 y_2 - u_2 v_2))$$

wobei gilt:

- B_i = Anfangsbonus in Phase i
- N_i = totale Anzahl Aufgaben in Phase i
- x_i = durchschnittlicher Anteil richtig gelöster Aufgaben in Phase i

y_i	=	durchschnittlicher Gewinn (nach Abzug der Ausgaben) bei richtig gelöster Aufgabe in Phase i
u_i	=	$(1 - x_i)$ = durchschnittlicher Anteil falsch gelöster Aufgaben in Phase i
v_1	=	$(1'000 - y_1)$ = durchschnittlicher Verlust bei falsch gelöster Aufgabe in Phase 1
v_2	=	$(1 - y_2)$ = durchschnittlicher Verlust bei falsch gelöster Aufgabe in Phase 2
D	=	Divisor

4.11 Abschliessende Bemerkungen

Der vorliegende Forschungsbericht dokumentiert die Weiterentwicklung eines in der aktuellen Entscheidungspsychologie gängigen Experimentaldesigns (Information Board) zur Untersuchung von Entscheidungsstrategien. Neu an diesem Design ist die Ausweitung auf den Mehralternativenfall. (In den bisherigen Experimenten hatten die Versuchspersonen jeweils die Wahl zwischen genau 2 Alternativen). Des Weiteren fokussieren die Versuchspersonen mit dem Design stärker auf das Lösen von Aufgaben (d.h. auf die Beantwortung der Quizfragen) und rücken damit den Fokus von der Informationssuche weg. Das Experiment verliert damit im positiven Sinne an Durchschaubarkeit.

Dank Process-Tracing (PT), d.h. der Aufzeichnung der Datenspuren während der experimentellen Datenerhebung, insbesondere während der Cuesuche, lässt sich das Informationssuchverhalten einer Versuchsperson mit verschiedenen theoretisch postulierten Suchstrategien vergleichen. Der eindeutige Rückschluss von den aufgezeichneten Datenspuren auf genau eine der postulierten Informationssuch-Strategien war bisher allerdings häufig unmöglich, weil sich die normativen Cuefolgen (Suchreihenfolgen) der postulierten Suchstrategien vielfach überlagerten - und dies nicht selten bereits von Beginn einer Cuefolge an. Das Problem war, dass die Validitäten und Diskriminationsraten der Cues voneinander abhängig und nicht kontinuierlich wählbar waren. Mit dem neuen Experimentaldesign (Joker) ist dieses Problem entschärft, da die Validitäten und Diskriminationsraten der Cues auf der Designebene erstens kontinuierlich und zweitens völlig unabhängig voneinander wählbar sind. Dies ist möglich, weil wir einerseits als Cues personalisierte Informanten verwenden und andererseits von Aufgabe zu Aufgabe alle Informanten durch neue ersetzen.

Bei den Informanten handelt es sich um virtuelle Experten (Telefonjoker), die zu einer Wissensfrage einen Antwort-Tipp abgeben. Dabei entspricht die Diskriminationsrate eines Cues der Wahrscheinlichkeit, mit der ein Experte telefonisch erreichbar ist und die Validität entspricht der Wahrscheinlichkeit, mit der die Antwort dieses Experten korrekt ist. Im Vergleich zu anderen Experimenten erreicht das neue Design somit eine deutlich grössere Alltagsnähe und damit eine bessere Anschaulichkeit der beiden Cuegüte-Konzepte Validität und Diskriminationsrate.

Für die adäquate empirische Überprüfung von Urteilsheuristiken muss sichergestellt sein, dass eine Versuchsperson für ihre Urteile ausschliesslich die im Experiment verwendete Informationsbasis verwendet, da sonst das Entscheidungsverhalten nicht valide gemessen werden kann. Bröder (2000a; 2004, S. 91) diskutiert dies als Identifizierungsproblem. Um den Einfluss von

Vorwissen ausschliessen zu können, wurden deshalb bisher häufig vollständig künstliche und damit alltagsferne Settings verwendet. In der Quizshow hingegen stellen wir den Versuchspersonen Almanachfragen, d.h. schwierige Wissensfragen, zu realen Sachverhalten aus den verschiedensten Wissensgebieten. Dabei ist entscheidend, dass wir trotz der Verwendung eines alltagsnahen Settings den Einfluss von Vorwissen über die Zielvariablen (d.h. die Bestimmung der korrekten Lösung) praktisch ausschliessen oder zumindest auf ein sehr tiefes Niveau hinunterdrücken können. Wir erreichen dies, indem wir die Versuchspersonen, gemessen am durchschnittlichen Alltagswissen, ausschliesslich schwierige Wissensfragen beantworten lassen. Sollte eine Versuchsperson bei einer Aufgabe (Frage) trotzdem Vorwissen zur Anwendung bringen können, so wird dies a) durch das Konfidenz-Rating identifiziert, baut b) die Versuchsperson motivational auf und wird c) bei der Messung der Suchstrategie durch eine ausreichend grosse Anzahl an Aufgaben aufgefangen.

Dasselbe gilt auch dann, wenn wir den Versuchspersonen zwischendurch mit Absicht leichtere bzw. leichte Fragen präsentieren. Dadurch, dass sie von Zeit zu Zeit von ihrem Vorwissen profitieren können, verhindern wir, dass die Versuchspersonen durch zu lange Serien unbeantwortbarer Fragen zu sehr frustriert werden und als Folge ihr Interesse am Experiment verlieren. Aufgaben, die aufgrund fehlender Informationssuche keinen Beitrag zur Klärung der Frage der verwendeten Informationssuch-Strategie liefern, können bei der Datenauswertung einfach ignoriert werden.

Als Computerprogramm ermöglicht die Quizshow mit minimalen Anpassungen auch die Implementation rein künstlicher Designs. Zudem können Validitäten und/oder Diskriminationsraten der Cues, falls gewünscht, stabil gehalten werden, indem den Versuchspersonen in jeder Aufgabe dieselben virtuellen Experten zur Verfügung gestellt werden. Diese Flexibilität ermöglicht einen Einsatz der Software für ein breiteres Spektrum an Fragestellungen zur Informationssuche beim Entscheiden.

4.12 Literatur

- Christen, S., Hausmann, D., & Läge, D. (2006). *Ist ein Gefühl für Informationsgüte lernbar? Ein Lernexperiment zu Validität und Diskriminationsrate von Informationsquellen*. AKZ-Forschungsbericht 37. Zürich: Angewandte Kognitionspsychologie.
- Brehmer, A., & Brehmer, B. (1988). What have we learned about human judgment from thirty years of policy capturing? In Brehmer, B. & Joyce, C.R.B. (Eds.), *Human Judgment: The SJT View*. Amsterdam: Elsevier Science.
- Brehmer, B. (1994). The psychology of linear judgement models. *Acta Psychologica*, 87, 137-154.
- Bröder, A. (2000a). Assessing the empirical validity of the "Take The Best"-heuristic as a model of human probabilistic inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 1332-1346.
- Bröder, A. (2000b). *"Take the best - Ignore the rest" Wann entscheiden Menschen begrenzt rational?* Lengerich: Pabst Science Publishers.
- Bröder, A. (2005). *Entscheiden mit der "Adaptiven Werkzeugkiste". Ein empirisches Forschungsprogramm*. Lengerich: Pabst Science Publishers.

- Bröder, A., & Schiffer, S. (2003b). Take The Best versus simultaneous feature matching: Probabilistic inferences from memory and effects of representation format. *Journal of Experimental Psychology: General*, 132, 277–293.
- Böckenholt, U., & Hynan, L. (1994a). Caveats on a process-tracing measure and a remedy. *Journal of Behavioral Decision Making*, 7, 103–118.
- Böckenholt, U., & Hynan, L. (1994b). Differences and similarities between SI and SM: A reply to Payne and Bettman. *Journal of Behavioral Decision Making*, 7, 123–128.
- Edbauer, M. (Hrsg.) (2001). *Wissensportal Allgemeinbildung. Trautwein Lexikon-Edition, Genehmigte Sonderausgabe*. München: Compact.
- Einhorn, H. J., Kleinmuntz, D. N., & Kleinmuntz, B. (1979). Linear regression and process-tracing models of judgment. *Psychological Review*, 86, 465–485.
- Fiedler, F.E. (1978). The contingency model and the dynamics of the leadership process. In L. Berkowitz (Ed.), *Advances in experimental social psychology*, 11, 59–112. New York: Academic Press.
- Gigerenzer, G., & Goldstein, D.G. (1996). Reasoning the fast and frugal way: models of bounded rationality. *Psychological Review*, 103, 650–669.
- Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review*, 98, 506–528.
- Harte, J. M. (1995). *Multiattribute evaluation processes: Evaluation types and research techniques*. Amsterdam: Universiteit van Amsterdam.
- Harte, J.M., & Koele, P. (2001). Modelling and describing human judgement processes: The multiattribute evaluation case. *Thinking & Reasoning*, 7, 29–49.
- Läge, D., Hausmann, D., & Christen, S. (2005). *Wie viel bezahlen für eine valide Information? Suchkosten als limitierender Faktor der Informationssuche*. AKZ-Forschungsbericht 07. Universität Zürich: Angewandte Kognitionspsychologie.
- Payne, J. W. (1976). Task complexity and contingent processing in decision making: An information search and protocol analysis. *Organizational Behavior and Human Performance*, 16, 366–387.
- Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14(3), 534–552.
- Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. Cambridge: Cambridge University Press.
- Stewart T.R. (1988). Judgment analysis: Procedures. In Brehmer, B., & Joyce, C.R.B. (Eds.), *Human Judgment: The SJT View*. Amsterdam: Elsevier Science.
- Sundström, G. (1987). Information search and decision making: the effects of information displays. *Acta Psychologica*, 65, 165–179.

Kapitel 5

Was macht einen „guten Cue“ aus? Strategien der Informationssuche beim heuristischen Entscheiden unter Unsicherheit

Im Rahmen des Konzepts der Ökologischen Rationalität wird angenommen, dass Entscheidungen in der Regel auf der Basis von zu suchender Information gefällt werden und dass diese Informationssuche Kosten verursacht. Die Take-the-Best-Heuristik (TTB) kann mit ihren drei building blocks von search rule, stopping rule und decision rule derartige kostenintensive Informationssuche und die darauf aufbauende Entscheidung modellieren. Die beiden vorgestellten Experimente messen alle drei Phasen dieses Modell. Während die Stopp- und Entscheidungsregel für weitgehendes One-Reason Decision-Making sprechen, entspricht die Suchregel nur in einer kleinen Zahl von Fällen der durch TTB vorhergesagten Validitätsfolge. Success (als Kombination von Validität und Diskriminationsrate) erweist sich gesamthaft als das bessere Modell. Die Auswertungen auf individueller Ebene zeigen jedoch, dass nur ganz wenige Personen überhaupt eine stabile Suchregel verfolgen.

5.1 Einleitung

Für gewöhnlich wird Entscheiden als ein kognitiver Prozess angesehen, bei dem eine gegebene Informationslage durch Schlussfolgerung in eine neue Informationslage oder in eine Handlungsabsicht transformiert wird. Umstritten sind dabei die Algorithmen, die für einen solchen Transformationsprozess herangezogen werden: Die psychologischen Modelle für menschliches Verhalten reichen von komplexen mathematischen Formeln (z.B. dem Bayes-Theorem zur Bestimmung von Eintretenswahrscheinlichkeiten oder der Regression zur Errechnung eines maximalen Nutzens) über die Identifikation von *Heuristics and Biases* – die vor allem von der Arbeit von Tversky & Kahneman (1974) inspiriert sind – bis hin zu einfachen Heuristiken, die den Maximen einer Ökologischen Rationalität folgen (z.B. Gigerenzer, Todd and the ABC Research Group, 1999).

Bettet man den Moment, in dem eine Entscheidung gefällt wird, allerdings in den Gesamtvollzug menschlichen Handelns ein, dann wird schnell klar, dass die Entscheidung selbst nur der Kulminationspunkt eines längerfristigen Prozesses ist, an deren Anfang ein Problem und an deren Ende eine Handlung zur Lösung dieses Problems steht. Wie Simon (1955; 1956; 1982) bereits beschrieben hat, geht dem Moment des Entscheidens eine Phase der Informationssuche voraus. Diese führt zu einer zumeist nicht vollständigen Informationslage, und an diese natürliche Situation sind, so Simon, die Algorithmen menschlicher Informationsverarbeitung angepasst. Im Gegensatz zu einer von idealistischen mathematischen Vorstellungen ausgehenden *unbounded rationality* bezeichnet er die menschliche Verarbeitungskapazität als *bounded rationality*. Die Grundidee besteht darin, dass Organismen gar nicht die denkbar beste Entscheidung zu fällen brauchen, sondern lediglich eine hinlänglich nützliche. Durch die Mechanismen der Evolution, so die Argumentation, haben sich Organismen gebildet, die für ihre Entscheidung nur ein gewisses Mass an Informationen suchen und verarbeiten können müssen, um zu diesen hinlänglich nützlichen Entscheidungen in der Lage zu sein. Ein wichtiger Faktor für die Entwicklung umweltgerechten Verhaltens ist die Tatsache, dass die notwendige Informationssuche Ressourcen (zumindest Zeit und Kalorien) kostet (vgl. auch Sargent, 1993).

Informationssuche wird in der experimentellen Entscheidungsforschung der vergangenen Jahrzehnte zumeist ausgeklammert (Arbeiten wie die von Connolly & Gilani, 1982; Payne, Bettman & Johnson, 1993; Saad & Russo, 1996 stellen eher die Ausnahme dar). Vielmehr wird im Standardexperiment eine bestimmte Informationslage angeboten, und die Versuchsperson hat lediglich den kognitiven Transformationsprozess zu erbringen (was häufig auf eine wenig optimale Art geschieht, wie die Forschung der Arbeitsgruppen um Kahneman und Tversky eindrucksvoll gezeigt hat).

5.1.1 Take The Best als ökologisch rationale Entscheidungsheuristik

Gigerenzer hat in den vergangenen Jahren eine Reihe von Entscheidungsmodellen formuliert, die Simons Argumentation der *bounded rationality* Rechnung tragen und dabei zumindest potentiell so gebaut sind, dass sie eine Phase der Informationssuche einschliessen (Gigerenzer, Hoffrage & Kleinbölting, 1991; Gigerenzer & Goldstein, 1996). Es handelt sich dabei um Verfahren des

One-Reason decision making, bei dem Informationen nur so lange (gesucht und) verarbeitet werden, bis eine ausreichende Basis für die Entscheidung vorliegt. Der Prototyp für solche Verfahren ist die *Take-the-Best-Ignore-the-Rest-Heuristik* (kurz TTB genannt). Nach ihr werden sukzessive die möglichen Informationsquellen in der Reihenfolge ihrer Güte befragt, und sobald eine dieser Quellen eine Entscheidungsgrundlage offenbart, wird die Informationsverarbeitung abgebrochen und lediglich auf der Basis dieser Informationsquelle die Entscheidung gefällt.

Der von TTB spezifizierte Entscheidungsweg sei anhand eines einfachen Beispiels erläutert: Zu entscheiden sei die Frage, welche von zwei deutschen Städten mehr Einwohner habe, Köln oder Hannover. Sind keine unmittelbaren Einwohnerzahlen verfügbar, so ist eine Entscheidung unter Unsicherheit zu fällen. Der Weg des TTB-Modells geht dabei über indirekte Hinweise auf die Zielvariable, sog. *Cue-Informationen*. Ein Cue (z.B. die Information, ob eine Stadt Kantonshauptstadt ist) indiziert mit einer gewissen Wahrscheinlichkeit die korrekte Richtung der Entscheidung (Kantonshauptstädte pflegen im Mittel grösser zu sein als Nicht-Kantonshauptstädte). Diese Wahrscheinlichkeit wird die *Validität* des Cues genannt. Da sie in der Regel kleiner als 100% ist, bekommt man zwar eine Wahrscheinlichkeitsaussage, aber keine Sicherheit. Trotzdem kann man sich an eine solche Information halten und trifft im Mittel viele korrekte Entscheidungen.

Eine Person, die nun also entscheiden soll, ob Köln oder Hannover mehr Einwohner hat, macht sich auf die Suche nach der bestmöglichen Information (diese Suche kann im Gedächtnis oder in der Welt stattfinden). Der beste Cue für deutsche Städte ist der Bundeshauptstadt-Cue (die Hauptstadt ist die grösste Stadt, d.h. der Cue hat eine Validität von 100%). Er trifft aber weder auf Köln noch auf Hannover zu, ermöglicht also keine *Diskrimination* zwischen den beiden Städten (die Diskriminationsrate ist beim Bundeshauptstadt-Cue extrem gering, da er nur bei einem winzigen Anteil von Städtepaaren unterscheidet, nämlich in denen Berlin vorkommt). Also sucht die Person weiter. Der zweitvalideste Cue ist der Messestadt-Cue (in 91% der Fälle ist eine deutsche Grossstadt mit Messezentrum grösser als eine ohne). Dieser Cue trifft sowohl auf Köln als auch auf Hannover zu, ermöglicht also ebenfalls keine Entscheidung. Der drittbeste Cue für deutsche Grossstädte ist der Bundesliga-Cue (Validität: 87%). Köln besitzt eine Fussballmannschaft in der Bundesliga, Hannover nicht. Die Person bricht an dieser Stelle die Informationssuche ab und entscheidet sich für Köln (womit sie richtig liegt, denn Köln hat ca. 400.000 Einwohner mehr als Hannover).

Gigerenzers TTB-Modell kann somit durch drei Regeln beschrieben werden:

1. *Suchregel*: Gehe die Cues nach diskriminierenden Informationen durch, beginnend beim besten Cue (das ist der Cue mit der höchsten Validität).
2. *Stoppregel*: Beende die Informationssuche, sobald Du einen diskriminierenden Cue gefunden hast.
3. *Entscheidungsregel*: Fülle die Entscheidung in Richtung des diskriminierenden Cues.

Damit liegt ein *nicht-kompensatorisches* Entscheidungsmodell vor (vgl. Martignon & Hoffrage, 1999): Wegen der Stopregel können Informationen von niedriger validen Cues – auch wenn sie alle in die Gegenrichtung zeigen – die Entscheidung nicht mehr umbiegen. Auf den ersten Blick mag ein solches Verfahren, das Gigerenzer selbst als „fast and frugal“ bezeichnet, wie ein „Notnagel“ aussehen: „Optimale“ Entscheide sollten ja so viel Information wie möglich in mathematisch akkurat gewichteter Weise berücksichtigen. Doch Simulationsstudien zeigen, dass dem nicht so ist: Die *Take-the-Best-Ignore-the-Rest-Heuristik* erzielt in Entscheidungssituationen im Mittel die gleiche Performanz wie aufwendige regressionsanalytische Modelle, die sämtliche verfügbare Information integrieren (Gigerenzer & Goldstein, 1996; Gigerenzer, Czerlinski & Martignon, 1999; Czerlinski, Gigerenzer & Goldstein, 1999).

5.1.2 Empirische Evidenz: Take The Best-Entscheidungen nur bei Informationskosten

Bröder (2000a, 2001) hat zu Recht darauf hingewiesen, dass es für ein überzeugendes psychologisches Modell allerdings nicht ausreicht, normativ-algebraisch zu zeigen, dass eine solche Heuristik zu gleich guten Entscheidungen führt wie eine mathematische Norm der *unbounded rationality*. Es müsse auch der Nachweis geführt werden, dass Menschen tatsächlich wie in diesem Modell beschriebenen Informationen suchen und für ihre Entscheidungen verarbeiten. Erste experimentelle Daten wurden von Rieskamp & Hoffrage (1999) vorgelegt. Sie variierten in einem *Information Board* Design die Variable der zur Verfügung stehenden Zeit und fanden, dass unter Zeitdruck (und nur dann) 23% ihrer Versuchspersonen eine Strategie verwenden, die sich als TTB-konform deklarieren liess. Von Kommentatoren dieser Publikation (Chater, 2000; Cooper, 2000; Oaksford, 2000) werden diese Befunde aber als zu dürftig kritisiert, um *Take The Best* als Modell für menschliches Entscheidungsverhalten akzeptieren zu können. Auch in ihrer Antwort auf Bröder (2001) konnten Hertwig & Hoffrage (2001b) keine weitere empirische Evidenz für TTB vorlegen, die über die Befunde von Rieskamp & Hoffrage (1999) hinausgingen.

Bröder (2000b) konnte in seinen Experimenten Zeitdruck nicht als eine Bedingung identifizieren, die zu vermehrtem Gebrauch von TTB führte. Vielmehr legt er Daten vor, die den Faktor der Informationskosten als notwendige Bedingung aufzeigen, damit Personen in substantiellem Umfang die TTB-Strategie verwenden (Experimente 3 und 4). In einem fiktiven Aktienspiel – ähnlich dem von Rieskamp & Hoffrage (1999) – stellt er seinen Versuchspersonen vier unterschiedlich valide Cues mit Informationen zur Verfügung, aufgrund derer aus zwei Aktien diejenige mit der wahrscheinlich besseren Performanz auszuwählen ist. Im Falle einer korrekten Vorhersage erhält die Versuchsperson einen Gewinn von 1000 Penunzen, aber sie muss für jeden Cue, den sie betrachten will, 100 Penunzen zahlen. Durch Einführung dieser hohen Informationskosten konnte Bröder bis zu 65% seiner Versuchspersonen hinsichtlich ihrer Entscheidungen als TTB-orientiert klassifizieren. Ohne entsprechende Informationskosten, so zeigen seine anderen Experimente, ist der Anteil TTB-klassifizierter Versuchspersonen äusserst gering.

Damit wurde zum ersten Mal *Take The Best* als ein gutes deskriptives Modell für tatsächliche Entscheidungen empirisch bestätigt. Allerdings verpasste Bröder die Chance, neben der Entscheidungsregel (die ja nur eine von drei Regeln in TTB ist) auch die Informationssuche in Form von Suchregel und Stopregel separiert zu berücksichtigen. Dieses Manko dürfte mit seiner

Grundentscheidung zusammenhängen, die Auswertung seiner Daten vollständig auf dem Zugang des *Structural Modeling* abzustützen. Damit modelliert er lediglich die Richtung der Entscheidung, nicht aber den Prozess der Informationssuche selbst. Bröder steht hier durchaus in der Tradition der Simulationsstudien der ABC-Arbeitsgruppe um Gigerenzer (z.B. Gigerenzer et al., 1999), die ebenfalls jeweils auf der Ebene einer gegebenen Informationslage einen Modellfit gerechnet haben. Soweit in diesen Simulationen Informationssuche vorkommt, handelt es sich um „Informationssuche im Gedächtnis“, aber nicht in der Welt.²⁴ Neben den weiterführenden Arbeiten von Bröder (2000a; 2000c; 2002; 2003; 2004; Bröder & Schiffer, 2003a) folgten weitere wichtige experimentelle Untersuchungen bezüglich TTB aus der Forschungsgruppe um Ben Newell und David Shanks: Newell & Shanks, 2003; Newell, Weston & Shanks, 2003; Newell, Rakow, Weston & Shanks, 2004. Bröders zog das Fazit, dass diese ersten experimentellen Untersuchungen deutlich machen, „dass die Verwendung der Take The Best-Heuristik offenbar sowohl von individuellen Präferenzen als auch von Kontextbedingungen abhängt.“ (Bröder, 2000a). Das Fazit aus diesen Studien zur Anwendung von TTB sieht denn auch ernüchternd aus, was Bröder (2004) mit den Worten „manche Personen verwenden die Heuristik manchmal“ zusammenfasste (S. 94).

5.1.3 Informationssuche und die Frage nach dem „guten Cue“

Um die drei Building Blocks von TTB überprüfen zu können, sind Datensätze über das tatsächliche Verhalten von Personen notwendig, in denen die Phase der Informationssuche als kritische Variable berücksichtigt ist. Neben der non-kompensatorischen Entscheidungsregel sind ja auch die Suchregel und die Stopregel in TTB eindeutig definiert: Gesucht wird in der Reihenfolge der Güte der Cues, und die Suche ist abubrechen, sobald der erste Cue diskriminiert.

In den bislang berichteten Simulationsstudien und Experimenten blieb stets die Frage nach dem „guten Cue“ unangetastet. Gigerenzer & Goldstein (1996) waren (in der Tradition des Ansatzes von Brunswik, 1955) davon ausgegangen, dass die Validität (V) eines Cues die massgebliche Grösse sei. Obwohl die Autoren zumindest mit *Take The Last* einen anderen Kandidaten benennen (bei dem Diskrimination in einer früheren Aufgabe eine Rolle spielt), blieben Alternativen zur Validität bislang sowohl in der Theoriebildung als auch in Simulationen und Experimenten völlig unberücksichtigt (eine der wenigen Ausnahmen ist in Gigerenzer et al., 1999, in Kapitel 6 dokumentiert). Bröder (2000b) beispielsweise veranstaltete in den Experimenten 1 und 2 ein aufwendiges Validitätstraining, um bei seinen Versuchspersonen eine Rangordnung der Güte der Cues zu implementieren. Die Theorie einer Ökologischen Rationalität nimmt nämlich an, dass die Validität von Cues für eine Referenzklasse durch viele Erfahrungen in der Vergangenheit implizit gelernt worden ist. Genauso plausibel erscheint dann aber, dass auch ein Wissen über die Zahl der Fälle erworben wurde, in denen eine Quelle überhaupt eine Information enthält, die zwischen den beiden Alternativen unterscheidet. Diese als *Diskriminationsrate* (D) bezeichnete

²⁴ Ein eindeutiges Indiz für die Modellierung einer reinen Informationssuche im Gedächtnis ist in einer Simulationsstudie jeweils die Berücksichtigung des Recognition-Prinzips. Informationssuche in der Welt kann auf diese Form von Cue nur in konstruierten Ausnahmefällen zugreifen.

Grösse wäre demnach also ebenfalls ein Hinweis auf die Güte eines Cues. Ferner liegen zwei Vorschläge für alternative Suchstrategien vor, welche V und D integrieren: Hoffrage, Hertwig & Czienskowski (2003) haben *Usefulness* als multiplikative Kombination eingeführt und Martignon & Hoffrage (1999; vgl. auch Martignon & Hoffrage, 2002) haben mit *Success* ein ähnliches Maß vorgeschlagen, welches Newell, Rakow, Weston & Shanks (2004) in ihren Lernexperimenten verwendet haben.

Für Experimente, die den Fokus besonders auf die Informationssuche legen, sollten daher mehrere Kandidaten zur Grundlage einer Verhaltensvorhersage bei der Informationssuche gemacht werden:

- *Validität (V)*: Eine Person sucht die Informationsangebote in der Reihenfolge der Validität ab, beginnend mit dem validesten Cue.
- *Diskriminationsrate (D)*: Eine Person sucht die Informationsangebote in der Reihenfolge der Diskriminationsrate ab, beginnend mit demjenigen Cue, der in den meisten Fällen zu einer Unterscheidung führt.
- *Usefulness (V*D)*: Eine Person versucht diese beiden Grössen zu integrieren, indem sie aus den zwei Werten das Produkt bildet und die Cues in der Reihenfolge der Grösse der resultierenden Produkte absucht, beginnend mit dem grössten Produkt. Die so resultierende Rangordnung entspricht dem Maß „Prozent korrekt“ und wurde unter als *Usefulness (U)* eingeführt (Hoffrage, Hertwig & Czienskowski, 2003).
- *Success (V*D + [1-D]*0.5)*: Diese Formel wird als *Success (S)* bezeichnet (Martignon & Hoffrage, 1999) und gibt den Wert eines Cues an, wenn keine Möglichkeit zur weiteren Suche besteht. Normativ sinnvoll ist dieses Mass eigentlich nur dann, wenn nach dem ersten Cue zu entscheiden ist, unabhängig davon ob er diskriminiert oder nicht. Dies ist im vorliegenden Experiment nicht der Fall, aber eine Versuchsperson könnte ja die genaue Implikation falsch verstehen. Deswegen wird dieses Mass als Hypothese mit aufgenommen, zumal Newell, Rakow, Weston & Shanks (2004) ihr Cuetraining und die kontinuierliche Informationssuche ebenfalls über *Success* spezifiziert haben (und nicht etwa über *Usefulness*, wie eigentlich in ihren Experimenten angebracht gewesen wäre).
- *Validität + Diskriminationsrate*: Ferner könnte eine Person intendieren, V und D gemeinsam zu berücksichtigen, jedoch nicht die Art des Zusammenhangs verstehen. Sie könnte demnach versuchen, diese beiden Grössen additiv zu integrieren, indem sie aus den zwei Werten die Summe bildet und die Cues in der Reihenfolge der Grösse der resultierenden Summe absucht, beginnend mit der grössten Summe.
- *PingPong-Strategie*: Eine Person kombiniert Validität und Diskriminationsrate, indem sie die Informationssuche entweder mit der höchsten Validität oder mit der höchsten Diskriminationsrate beginnt und dann kreuzweise mit dem anderen höchsten Wert weitermacht.

Jede dieser Suchstrategien ist mit einem One-Reason Decision Making (ORDM) vereinbar, auch wenn die *Take The Best*-Heuristik (TTB) einzig die Validität spezifiziert. Gigerenzer, Czerlinski & Martignon (1999) sowie Todd & Gigerenzer (2000) betonen, dass Menschen nicht nur über eine einzige Strategie verfügen, sondern über eine ganze *Adaptive Toolbox* (an Urteilsheuristiken), aus der sie ihr kognitives Vorgehen fallweise zusammensetzen. Theorien der Ökologischen Rationalität hätten deswegen *building blocks* zu identifizieren (im Falle von TTB sind das die *step-by-step procedures*, die *simple stopping rules* sowie das *One-Reason Decision Making*), innerhalb derer verschiedene Personen in unterschiedlichen Situationen bestimmte Werkzeuge einsetzen. Zwar sprengt eine Erweiterung der Kandidaten die Suchregel von *Take-the-Best*, doch würde es für eine Stützung der Heuristik als ganze ausreichen, wenn sich in homogen bleibenden Problemstellungen bei untersuchten Personen zumindest individuell stabile Vorlieben für das eine oder das andere Suchverfahren ergeben.

Somit sind zwei Fragestellungen zu beantworten:

1. Entspricht, wie in der traditionellen Formulierung von TTB angenommen, die Rangordnung der Güte der Cues der Rangordnung der Validität, wenn die Suche nach diesen Cues mit Kosten verbunden sind?
2. Haben Personen individuell stabile Informationssuchstrategien?

Zur Beantwortung dieser Fragen werden im Folgenden zwei Experimente berichtet. Das erste basiert auf den Datensätzen, die Gigerenzer & Goldstein (1996) dienten, um *Take The Best* als Heuristik zu beschreiben und in einer Computersimulation auf die Güte hin zu überprüfen. Diese Daten haben allerdings den Nachteil, dass unterschiedliche Erklärungsmodelle weitgehend identische Vorhersagen machen. Das gilt nicht nur für die Entscheidungsregel (wie bereits Bröder, 2000b zu Recht angemerkt hat), sondern in ähnlichem Ausmass auch für die Suchregel. Deswegen arbeiten wir in Experiment 2 mit einem Datensatz, der die verschiedenen oben vorgeschlagenen Möglichkeiten für Suchreihenfolgen maximal separiert und auf diese Weise individuell stabile Informationssuchstrategien zu identifizieren in der Lage ist.

Tabelle 5.1. zeigt das Beispiel zweier Cues mit unterschiedlicher Validität und Diskriminationsrate und erläutert damit die Verhaltensvorhersage der oben angeführten Modelle für die bevorzugte Suchreihenfolge. Dieses einfache Beispiel zeigt auch, dass sich durch die begrenzte Zahl der Freiheitsgrade die Vorhersagen der unterschiedlichen Kandidaten im Normalfall überlappen.

Tabelle 5.1: Zwei Cues mit unterschiedlicher Validität und Diskriminationsrate. Durch die begrenzte Zahl der Freiheitsgrade überlappen die Vorhersagen der unterschiedlichen Kandidaten.

	Cue A	Cue B
Diskriminationsrate (D)	.350	.510
Validität (V)	.900	.710
Usefulness (D*V)	.315	.362
Success (D*V+[1-D]*0.5)	.640	.607
D+V	1.250	1.220

5.2 Experiment 1

5.2.1 Aufbau des Experiments

Um die Bedeutung verschiedener Kandidaten für die Frage nach dem „guten Cue“ möglichst adäquat untersuchen zu können, sollten Validität und Diskriminationsrate nicht positiv miteinander korreliert sein. Gigerenzer & Goldstein (1996) berichten für ihre Daten eine leicht negative Korrelation zwischen diesen beiden Grössen, so dass sich das folgende Experiment an ihrem Material orientiert. Es fragt demzufolge eine Person, welche von zwei Städten die höhere Einwohnerzahl hat. Diejenigen 72 der 83 deutschen Städte wurden aus den Originaldaten von Gigerenzer & Goldstein ausgewählt, bei denen sich zum Zeitpunkt des Experiments keine der Cue-Informationen geändert hatte. Sie wurden nach ihrer Einwohnerzahl in drei Drittel geteilt, und innerhalb jedes Drittels wurden die Städte zufällig zu Paaren zusammengestellt. Damit ergaben sich 36 Paare, bei denen (durch die verhältnismässig ähnliche Einwohnerzahl) nie alle Cues diskriminierten. Keine Stadt kam in mehr als einem Paar vor, so dass Cue-Lernen innerhalb des Experiments nicht möglich war.

Tabelle 5.2: Liste der verwendeten Cues mit Validitäten und Diskriminationsraten (Übersetzung ins Englische)

No.	Chinese cue names	Cues used by Gigerenzer & Goldstein (1996)	Ecological validity	Discriminationrate
1	provincial capital	national capital	1.00	0.02
2	hospital	exposition site	0.91	0.22
3	volleyball team	soccer team	0.90	0.35
4	party office	IC train station	0.78	0.41
5	daily newspaper	state capital	0.77	0.26
6	public bus system	license plate	0.75	0.27
7	university	university	0.71	0.51
8	bicycle factory	industrial belt	0.56	0.30
9	statue of Mao	Eastern Germany	0.51	0.27

Ebenfalls zufällig wurden für jedes Paar vier der neun Cues aus Gigerenzer & Goldstein als Informationsangebot gezogen (Tabelle 5.2). Die Ziehung unterlag den Vorgaben, dass außer dem Bundeshauptstadt-Cue (der nur zweimal vorkam) jeder Cue in den 36 Aufgaben gleich oft verwendet wurde (die beiden Cues mit den Validitäten .91 und .90 17mal, alle anderen 18mal) und dass mindestens 1 Cue zwischen den beiden Städten eines Paares diskriminierte²⁵.

Um für die Versuchspersonen jegliches Vorwissen auszuschliessen, wurden die deutschen Städtenamen gegen fiktive chinesische ausgetauscht, genauso wie die Namen der Cues für deutsche

²⁵ Die Validität zweier Cues wurde gegenüber Gigerenzer & Goldstein ein wenig angepasst (soccer team 0.87 nach 0.90 sowie License Plate 0.75 nach 0.74).

Städte durch „chinatypische“ Cues gleicher Validität und Diskriminationsrate ersetzt sind (Tabelle 5.2). Dieses Setting mit fiktiven Informationen garantiert, dass der Wissensstand der Versuchsperson für die Auswertung der Daten vollständig bekannt ist.

Auf einem *Information Board* erhielt die Versuchsperson jeweils ein Städtepaar sowie ein Angebot von vier Cues präsentiert. Aus diesen vier Cues konnte die Person nun sequentiell Informationen „kaufen“, wobei bei jedem Cue Validität und Diskriminationsrate angegeben waren (die beiden Konzepte wurden mit den normalsprachlichen deutschen Begriffen „Trefferrate“ und „Anwendbarkeit“ bezeichnet und den Versuchspersonen in der Instruktion erläutert). Aufgrund eines sofortigen Feedback über die Cue-Information war sie in der Lage, die Strategie der weiteren Informationssuche, den Abbruch des Suchprozesses sowie die Richtung der Entscheidung zu bestimmen.

Nachdem die ersten Experimente zu Take The Best immer zuerst ein Validitätstraining absolvieren liessen und dabei zeigten, dass die Güte von Cues im Grundsatz durchaus lernbar ist (z.B. Bröder, 2000b), ist man inzwischen dazu übergegangen, wegen der höheren Kontrollierbarkeit des „Wissens“ der Versuchsperson und zur Abkürzung des experimentellen Verfahrens die Cuegüte vorzugeben (z.B. Bröder, 2000b; 2004). Das vorliegende Experiment ist diesem Weg ebenfalls gefolgt und hat Validität und Diskriminationsrate der verwendeten Cues jeweils für die Versuchspersonen gut sichtbar angegeben.

Die Versuchsperson erhielt die fiktive Aufgabe, für ein europäisches Unternehmen in den Städten einer (genauso fiktiven) chinesischen Provinz Verkaufsfilialen zu eröffnen. Dabei erhöht sich der Unternehmensgewinn, wenn sie jeweils aus einem Paar von Städten die grössere als Filialniederlassung wählt. Angaben über Einwohnerzahlen existieren keine, jedoch kann die Versuchsperson gegen ein Bestechungsgeld von einem chinesischen Beamten landestypische Cue-Informationen erhalten, die dafür sprechen, dass eine von beiden Städten die größere ist.

Das Bestechungsgeld betrug \$ 1'000 pro gekauftem Cue, und pro richtiger Entscheidung erhöhte die Versuchsperson den Unternehmensgewinn um \$ 10'000. Jede Versuchsperson wurde darüber informiert, dass von den 30 Teilnehmern des Experiments das Drittel mit den besten Gesamtergebnissen an einer Verlosung teilnehmen würde. Zu gewinnen waren 3 x 100 Schweizer Franken. Damit sollten die Versuchspersonen motiviert werden, möglichst gute Resultate zu erzielen, ohne aber in eine Extremstrategie zu verfallen.

Die 30 Versuchspersonen wurden aus der erwachsenen Schweizer Bevölkerung ausgewählt (Alter von 18 bis 57, Mittelwert: 35 Jahre), wobei keine Studierende der Psychologie oder der Wirtschaftswissenschaften zugelassen waren. Das Geschlecht war in etwa gleichverteilt (13 Frauen, 17 Männer), das Experiment fand an einem Notebook und in möglichst natürlicher Umgebung statt (bei den Versuchspersonen zu Hause oder am Arbeitsplatz), so dass der Charakter des Laborexperimentes vermieden wurde.

5.2.2 Resultate

Stoppregel und Entscheidungsregel. Die 30 Respondenten bearbeiteten insgesamt 1080 Aufgaben (siehe Abbildung 5.1). Dabei entsprachen sie in 768 Fällen der Stoppregel von ORDM, d.h. sie beendeten die Suche exakt nach dem Finden des ersten diskriminierenden Cues (36 mal wurde geraten, 54 mal ohne diskriminierenden Cue abgebrochen und 222 mal nach einem diskriminierenden Cue weitergesucht). Mit knapp 74% korrekter Vorhersagen schlägt sich die TTB-Stoppregel also nicht schlecht. Fast perfekt ist die Entscheidungsregel: Nur in 9 der 768 Fälle, die der Stopp-Regel entsprachen, entschied sich die Versuchsperson gegen die Richtung des diskriminierenden Cues. Das ist eine Vorhersagequote von beinahe 99%.

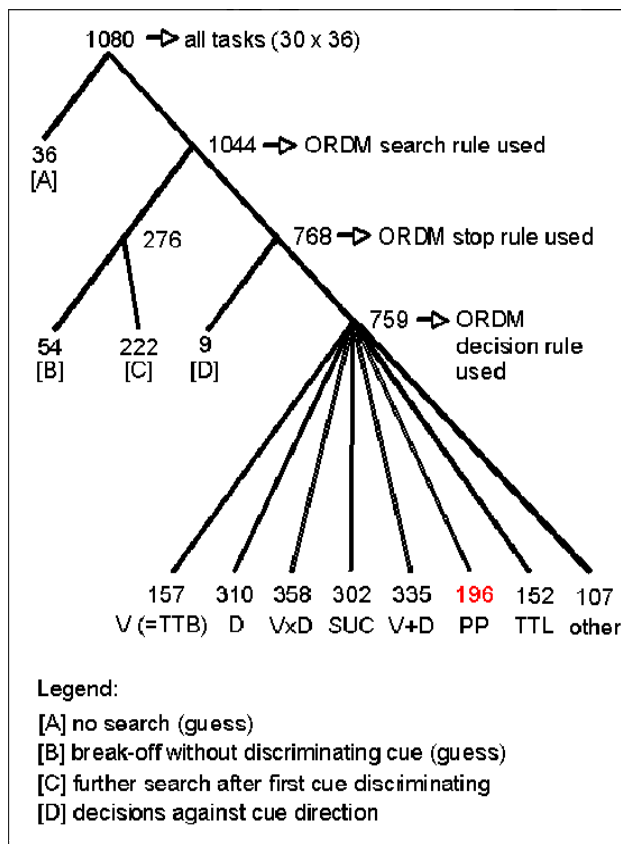


Abbildung 5.1: Anzahl der Fälle, die sich perfekt durch TTB beschreiben lassen. Die Fälle sind primär nach einem One-Reason Decision Making (ORDM) ausgerichtet, also nach der korrekt angewendeten Stoppregel (Anzahl Fälle, in denen die Suche nach dem ersten diskriminierenden Cue-Hinweis beendetet wurde). In insgesamt 759 Aufgaben wurde zudem die Entscheidungsregel von ORDM korrekt angewendet. Auf diese Zahl beziehen sich die untersuchten Suchstrategien (V=TTB, D, D*V, Suc, V+D, PP, TTL und übrige), wobei sich die einzelnen Strategien hochgradig überlappen.

Suchregel (Stichprobenebene). Für die gerade erfolgte Überprüfung der Stopp- und Entscheidungsregel wurde lediglich ORDM überprüft, aber nicht die Reihenfolge in der Informationssuche. Diese spezifiziert TTB in seiner klassischen Version als Validitätsfolge. Von den 1044 Aufgaben, bei denen die Versuchspersonen überhaupt nach Information gesucht haben, entsprachen 157 der Reihenfolge, die sich durch die Validität ergibt. 310 Aufgaben mit Informationssuche werden durch die Folge der Diskriminationsraten korrekt beschrieben, 358 durch Usefulness, 302 durch Success, 335 durch V+D und 493 durch eine Ping-Pong-Strategie (wovon in mehr als zwei Drittel der Fälle zuerst der Cue mit der höchsten Diskriminationsrate gekauft wurde, und wenn dieser nicht diskriminierte, der verbleibende Cue mit der grössten Validität). Interessant

ist, dass die Hypothese, allein die Diskriminationsrate bestimme den Cue-Kauf, mit 310 gegenüber 157 Fällen die Validität klar aussticht.

Suchregel (Ebene der Individuen). Betrachtet man dieselben Ergebnisse auf der Ebene der einzelnen Individuen (Tabelle 5.3), dann sieht man, dass sich bei 20 der 30 Versuchspersonen mindestens zwei Drittel aller Lösungen in Informationssuche *und* Entscheidung perfekt durch ein ORDM erklären lassen.

Tabelle 5.3 schlüsselt weiterhin die 30 Personen, die gemäss aller drei *building blocks* von TTBSuchen und entscheiden, hinsichtlich ihrer Suchstrategien auf. *PingPong* macht bei 17 Personen die besten Vorhersagen, die Diskriminationsrate und die multiplikative Verknüpfung von V und D ist für jeweils eine Person das beste Modell. Die Validität alleine spielt ebenfalls nur für eine einzige Person die entscheidende Rolle bei der Suche nach Information.

Tabelle 5.3: Aufschlüsselung der drei *building blocks* (Suchregel, Stoppregel und Entscheidungsregel) nach einzelnen Versuchspersonen (36 Aufgaben pro Versuchsperson)

Vp-Nr.	ORDM Regeln			Cue Such-Strategien						
	gesuchte cues	Stoppregel	Entsch.-regel	V	D	VxD	SUC	V+D	PP	other
4	36	36	36	12	6	11	29	22	5	8
13	36	36	34	9	8	13	14	16	3	6
16	36	36	36	4	20	19	10	12	1	5
22	36	36	36	7	13	13	13	11	3	4
23	36	36	35	3	24	25	8	14	2	4
29	36	36	36	6	16	15	17	21	4	10
5	36	35	35	1	34	25	6	14	4	2
11	36	33	33	2	10	10	7	9	3	1
2	36	32	32	14	7	9	19	18	6	6
7	36	32	32	14	4	6	25	18	7	5
21	36	32	32	2	18	20	6	12	1	2
25	36	31	31	2	21	21	9	13	3	4
8	36	30	30	12	5	10	16	12	7	4
24	36	30	30	4	10	11	8	12	3	4
26	36	30	30	8	14	19	12	12	4	1
10	35	29	27	1	16	14	4	9	4	1
14	36	28	28	5	17	21	14	18	0	8
17	36	28	28	7	10	12	13	13	4	4
6	36	27	26	3	10	12	5	8	4	4
27	36	25	25	10	1	5	9	5	4	2
3	36	22	22	3	10	15	10	12	1	10
18	36	21	21	6	3	8	15	14	4	6
1	35	20	20	1	18	18	8	14	0	3
28	36	15	13	9	0	2	3	3	1	3
19	36	14	14	4	1	3	6	6	3	4
30	20	11	11	2	7	8	7	7	0	0
12	36	9	9	1	0	3	1	1	0	1
9	36	8	8	1	4	7	1	2	0	3
20	18	7	7	4	3	3	7	7	0	0
15	36	3	2	0	0	0	0	0	0	0

Suchregel (nach Cues). Drei der neun Cues wurden in mehr als 50% der Fälle gekauft. Es sind dies university (81,5%), volleyball team (79,4%) und party office (70,2%). Interessanterweise sind diese auch die drei Cues mit der höchsten Diskriminationsrate. Die beiden validesten Cues (provincial capital und hospital) liegen in der Kaufhäufigkeit nur auf den Plätzen 4 und 7.

5.2.3 Diskussion

Take The Best (TTB) wird in der Stoppregel und in der Entscheidungsregel weitgehend bestätigt: Die Personen betreiben ein One-Reason Decision Making (ORDM). Die TTB-Suchregel (in der Reihenfolge absteigender Cue-Validitäten) befolgen die Versuchspersonen hingegen nicht. Wie bereits angesprochen, machen die Erklärungsmodelle für die Informationssuche weitgehend überlappende Vorhersagen. Besonders Ping-Pong ist in dieser Weise sehr „flexibel“. Die Überlappungen werden auch dadurch verursacht, dass in vielen Fällen die Versuchsperson die Informationssuche bereits nach dem ersten Cue beenden kann, so dass die Variationsbreite der möglichen Reihenfolgen eng begrenzt ist. Deswegen kann nicht genau gesagt werden, nach welcher Suchregel eigentlich im Wesentlichen vorgegangen wurde. Klar ist nur: An der Validität allein orientieren sich die Versuchspersonen nicht, und sie haben auch kein individuell durchgehend angewendetes numerisches Modell, welches die Suchreihenfolge durch eine Kombination aus Validität und Diskriminationsrate erklärbar machen würde. (Auch *Take The Last* [Gigerenzer & Goldstein, 1996], bei dem der in der vorausgehenden Aufgabe erfolgreiche Cue verwendet wird, scheidet als Alternativerklärung aus. TTL wäre in 490 Fällen anwendbar gewesen, erklärte die Informationssuche aber nur 142-mal und bei keiner einzigen Person übermäßig häufig.)

5.3 Experiment 2

In Experiment 2 soll die Cue-Struktur so aufgebaut werden, dass Überlappungen zwischen den Modellen minimiert werden und damit die Grundlage geschaffen wird, um die von einer Person verwendete Suchstrategie bestmöglich zu isolieren. In diesem Experiment werden die möglichen Suchstrategien auf individueller Versuchspersonenebene sozusagen einem Härte-test unterzogen. In einer ersten Phase werden die fünf Erfolg versprechendsten Suchstrategien aus Experiment 1 (Validität (V), Diskriminationsrate (D), Usefulness ($V \cdot D$), Success ($V \cdot D + (1-D) \cdot 0.5$) und $V+D$) in mehreren Aufgaben paarweise gegeneinander getestet. Die beiden am häufigsten angewendeten Suchstrategien werden in einer zweiten Phase nochmals gegeneinander getestet. Damit soll der einerseits der Nachweis erbracht werden, welche der Suchstrategien Versuchspersonen in Urteilsaufgaben tatsächlich präferieren und andererseits wie konsistent eine solche Präferenzstrategie angewendet wird.

Die uns interessierenden Fragestellungen lauteten wie folgt:

Fragestellung 1: Gibt es Personen, die bevorzugt eine der fünf theoretisch postulierten Suchstrategien (V, D, $V+D$, $V \cdot D$, SUC) mehr oder weniger konsistent anwenden?

Fragestellung 2: Welche der fünf Suchstrategien werden bevorzugt angewendet?

Aus diesen Fragestellungen leiteten wir die folgenden Hypothesen ab:

- Hypothese 1: Eine der fünf theoretisch postulierten Suchstrategien wird von der Mehrzahl der Versuchspersonen bei der Informationssuche bevorzugt angewendet ("starke Hypothese").
- Hypothese 2: Es gibt einige Versuchspersonen die eine der fünf theoretisch postulierten Suchstrategien in der Mehrzahl der von ihnen bearbeiteten Aufgaben anwenden ("schwache Hypothese").
- Hypothese 3: Es werden eher Suchstrategien angewendet, welche V und D miteinander kombinieren, d.h. SUC, $V \cdot D$ oder $V + D$.

5.3.1 Aufbau des Experiments

Stichprobe: Am Experiment nahmen insgesamt 24 studentische Versuchspersonen (Vpn) teil (9 Männer und 15 Frauen). Vier dieser 24 Vpn mussten wir nachträglich von der weiteren Auswertung ausschliessen, weil sie die technisch notwendige minimale Anzahl auswertbarer Informationssuch-Schritte (Cue-Käufe) zur zuverlässigen Detektion der Suchstrategien unterschritten. Das Durchschnittsalter der verbliebenen 20 Versuchspersonen (8 Männer und 12 Frauen) lag bei 30.0 Jahren (range: 22 – 44. $SD = 5.7$).

Die Aufgabe der Versuchspersonen bestand darin, 70 schwierige Wissensfragen²⁶ aus unterschiedlichen Wissensgebieten richtig zu beantworten, mit dem Ziel, über alle 70 Aufgaben einen möglichst hohen Spielgewinn zu erreichen. Für jede richtig beantwortete Frage wurde ein konstanter Gewinnbetrag gutgeschrieben. Falsch beantwortete Fragen hatten keinen Abzug zur Folge. Zu jeder Frage boten wir vier Antwort-Alternativen an, von denen genau eine die richtige Antwort darstellte.

Nach der erstmaligen Präsentation einer (zufällig aus einem Pool gezogenen) Wissensfrage, musste eine Vp zunächst die von ihr präferierte Antwort-Alternative bezeichnen und darüber ein Konfidenzurteil abgeben. Danach folgte die Möglichkeit, Telefonjoker aktiv um ihre Tipps zu befragen (Informationssuche)²⁷. Pro Aufgabe standen maximal sechs Telefonjoker (Cues) zur Verfügung, welche mit Angabe ihrer Validität (V) und Diskriminationsrate (D) auf einem Information Board präsentiert wurden. Der aufgedeckte Tipp (sofern der Cue denn diskriminierte) bestand aus dem Hinweis auf eine der vier Antwortalternativen. Im gesamten Spielverlauf traf

²⁶ Sämtliche Wissensfragen und die zugehörigen Antwort-Alternativen haben wir unverändert dem Buch "Wissensportal Allgemeinbildung" (Edbauer, 2001) entnommen. In einem Vorversuch hatten wir 600 dieser Wissensfragen auf ihre Tauglichkeit für das Experiment hin überprüft, d.h. auf ihren Schwierigkeitsgrad und ihre Verständlichkeit hin getestet. Im definitiv verwendeten Fragepool verblieben 140 Fragen.

²⁷ Die Auswertung der bei allen Antworten erhobenen Antwortsicherheiten (Konfidenzen) ergab, dass die Vpn nur in 2.8 % aller Aufgaben in der Lage waren, die richtige Antwort bei einer subjektiven Konfidenz von über 90 % und ohne Einsatz von Telefonjokern korrekt anzugeben.

die Vpn nie auf zwei identische Cues²⁸. Um das Aufdecken sämtlicher Tipps zu vermeiden, wurden Suchkosten eingeführt (1/10 des möglichen Gewinns pro aufgedecktem Hinweis), welche sowohl bei korrekter als auch bei falscher Antwort vom Spielkonto abgezogen wurden²⁹. Die Vp bestätigte den individuellen Abbruchzeitpunkt der Cue-Abfrage und gab ein weiteres Konfidenzsurteil ab. Danach wurde ihr die korrekte Lösung der Quizfrage mit einem kurzen Erläuterungstext präsentiert und die Vp konnte zur nächsten Frage übergehen.

Nach einer kurzen Testphase folgten die 50 Aufgaben der Phase 1, welche die fünf postulierten Suchstrategien (V, D, V+D, V*D, SUC) paar- oder gruppenweise gegeneinander testete. Die Cue-Sets (Cue-Validitäten und Diskriminationsraten, sowie die diskriminierenden Ausprägungen von jeweils sechs Cues innerhalb einer Aufgabe) wurden so gesetzt, dass sich für mindestens zwei Suchstrategien eine unterschiedliche bzw. eindeutige Aufdeckreihenfolge ergab³⁰. Mit den insgesamt 50 Cue-Sets der Phase-1 konnte so jede der fünf Strategien genau 15mal in der Reihenfolge der ersten beiden Cue-Wahlen (meistens schon beim Aufdecken des ersten Cues) eindeutig detektieren bzw. die Anwendung einer der anderen Suchstrategien eindeutig ausschliessen. Bei der Konstruktion der Cue-Sets haben wir darauf geachtet, dass diese den folgenden Anforderungen genügen:

1. Jede Suchstrategie muss genügend oft eindeutig detektiert werden können (Genügend grosses N).
2. Jede Suchstrategie muss gleich oft eindeutig detektiert werden können (Fairness).
3. Die mit einem bestimmten Cue-Set detektierbaren Suchstrategien müssen spätestens mit der zweiten Strategie-konformen Cue-Wahl eindeutig detektiert d.h. von allen anderen Strategien unterschieden werden können
4. Alle Cues innerhalb eines Cue-Sets müssen sich – paarweise miteinander vergleichen – bezüglich ihrer Güte um einen minimalen Betrag d voneinander unterscheiden (V-, D-Strategie: $d = .01$; alle anderen kombinierten Strategien: $d = .03$)
5. Die Validitäten und Diskriminationsraten müssen zufallsverteilt sein und jede V-D-Kombination kommt maximal einmal im gesamten Experiment vor (damit kein Validitäts-Lernen stattfinden kann).

Aufgrund der ersten beiden Anforderungen mussten wir bereits vor Beginn des Experiments festlegen, welche Cues eines Cue-Sets diskriminieren und welche nicht. Die Cue-Werte (Joker-

²⁸ Damit konnte sichergestellt werden, dass die Cues völlig unabhängig voneinander waren und dass kein "Validitäts-Learning" stattfinden konnte.

²⁹ Die Versuchspersonen erhielten am Ende des Experiments einen realen Geldbetrag (ihren Spielgewinn) bar ausbezahlt. Der Spielgewinn hing in multiplikativer Weise von den in den beiden Phasen erspielten Beträgen ab. Beide Phasen waren somit gleich wichtig. Der maximal erreichbare Spielgewinn betrug CHF 66.-. Der im Mittel ausbezahlte Gewinn war jedoch deutlich kleiner (M CHF = 14.00; SD CHF = 5.54).

³⁰ Es zeigte sich, dass wir mit einer einzigen Aufgabe nur maximal drei verschiedene Suchstrategien innerhalb der ersten beiden Cue-Käufe voneinander unterscheiden können (gewöhnlich waren es sogar bloss eine oder zwei).

Tipps) wurden hingegen erst während des Versuchsablaufs, unter Berücksichtigung der Zuverlässigkeit (Validität) und der Erreichbarkeit (Diskriminationsrate) des Jokers vom Programm bestimmt.

Grössere Deltas (d ; siehe Anforderung 4) führen zu einer potentiell besseren Differenzierung der Cue-Güte innerhalb eines Cue-Sets. Damit erhöhten wir die Wahrscheinlichkeit, dass unter ähnlich guten Cues der bezüglich der verwendeten Suchstrategie objektiv bessere Cue als solcher erkannt und gewählt wird. Gleichzeitig sinkt dadurch die Wahrscheinlichkeit, dass Cues innerhalb der individuellen Such-Reihenfolge aufgrund von Berechnungsfehlern miteinander vertauscht werden und damit eine saubere Strategie-Detektion verunmöglichen. Die Anwendung einer bestimmten Suchstrategie konnte nur dann detektiert werden, wenn die Vpn die Cues in der exakt Strategie-konformen Reihenfolge aufdeckten. Damit legten wir folglich ein sehr strenges Kriterium an die Anwendung bzw. an die Detektion der Suchstrategien an!

Für Phase 2 stellten wir ebenfalls einen aus 50 Cue-Sets bestehenden Pool zusammen. Diesmal so, dass wir die Möglichkeit hatten, daraus für jedes beliebige Strategie-Paar genau 10 Cue-Sets pro Strategie auszuwählen, mit denen wir genau diese Strategie eindeutig detektieren konnten. Am Ende von Phase 1 mussten wir bestimmen, welche beiden Suchstrategien eine Vp am stärksten präferiert, um sie in Phase 2 noch einmal gegeneinander testen zu können. Ausgewählt wurden jeweils die beiden von der Versuchsperson in Phase 1 am häufigsten angewendeten Strategien. Im Falle eines Gleichstandes von zwei oder mehreren Strategien half ein Tiebreaker (mittels eines zusätzlichen Kriteriums) bei der Bestimmung. Der Tiebreaker kam insgesamt bei lediglich 5 Versuchspersonen zum Einsatz.

Die Reihenfolge der Cue-Sets über die Aufgaben und die Anordnung der Cues innerhalb einer Aufgabe wurden individuell variiert. Die beiden am häufigsten angewendeten Suchstrategien wurden in weiteren 20 Aufgaben direkt gegeneinander getestet. Der „Sieger“ wurde als die präferierte Suchstrategie einer Vp bezeichnet.

5.3.2 Resultate

Auswertung 1: Hypothesengeleitete Strategiedetektion

Tabelle 5.4 zeigt die Verteilung der detektierten Suchstrategien in Phase 1 und 2. Die Versuchspersonen orientieren sich bei der Informationssuche an unterschiedlichen Suchstrategien. Hypothese 1 (die starke Hypothese), welche postuliert, dass eine der fünf Suchstrategien von der Mehrzahl der Vpn bei der Informationssuche bevorzugt angewendet wird, muss somit eindeutig verworfen werden.

Die (schwache) Hypothese 2 kann hingegen bestätigt werden. Personen orientieren sich an einer der getesteten Suchstrategien, wobei die beiden beliebtesten Suchstrategien offensichtlich die Validitäts- und die Success-Strategie sind. Allerdings: Keine der fünf Suchstrategien wird von irgendeiner der Vpn durchgängig angewendet!

Tabelle 5.4: Strategie-Verteilung in Phase-1 und Phase-2 gemäss Auswertung 1

Vp.-Nr.	Phase-1					Phase-2					Strategie
	V	D	V+D	V*D	SUC	V	D	V+D	V*D	SUC	
1	2	2				2	0				V
2	6				1	4				0	V
3	2	1		1	2	2				3	SUC
4	8		1		4	5				1	V
5	9	1	1		4	7				4	V
6	2	1	3	11	9				4	9	SUC
7	1	2	1	2	1		2		3		V*D
8			2	2	4				0	5	SUC
9	2		5	1	6			1		6	SUC
11	2		1	4	3				1	2	SUC
12	4		2	1	5	2				4	SUC
15	1					2	0				V
16				6	8				7	7	SUC
17	1		3	3	10				0	6	SUC
18	1			2	1	1			1		V*D
19	2		2	2	5	5				1	V
21	2			6	2	0			4		V*D
22	5		1	1		5			1		V
23	5					2	0				V
24	1	1	2	3				0	1	SUC	

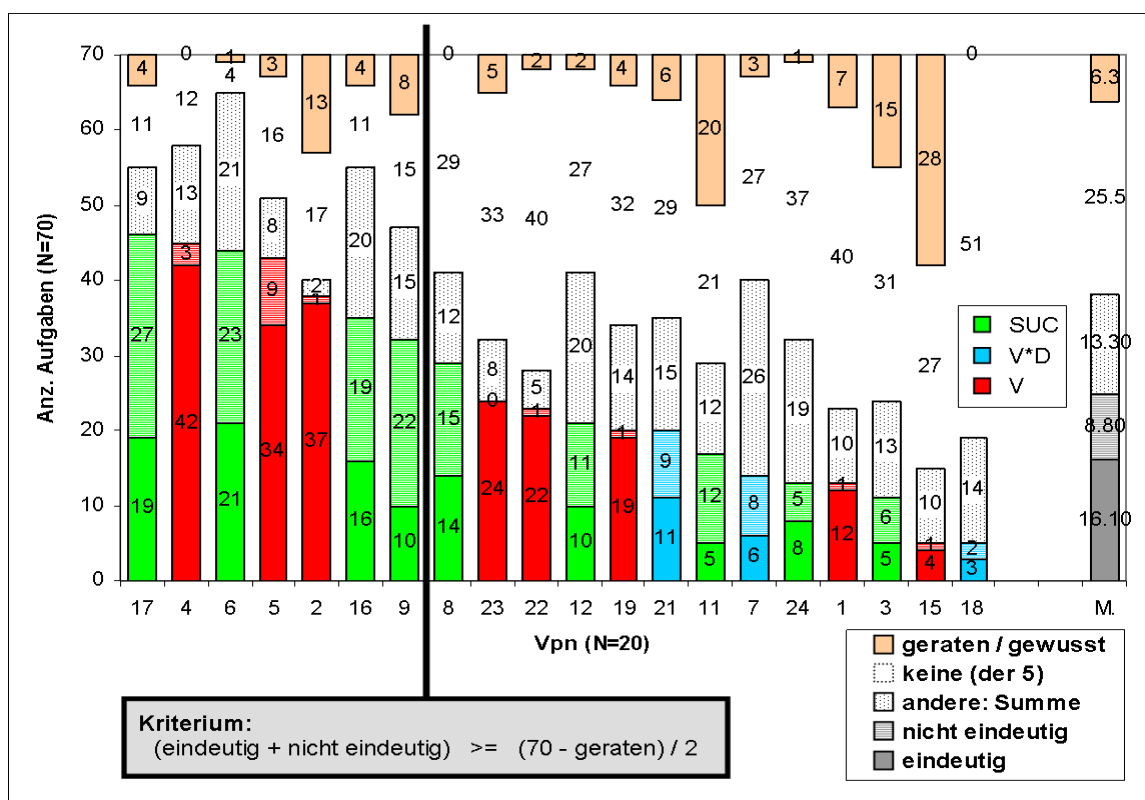


Abbildung 5.2: Strategie-Verteilung gemäss Auswertung 2

Hypothese 3 ("Es werden eher Suchstrategien angewendet, welche V und D miteinander kombinieren, d.h. SUC, V*D und V+D") kann insofern bestätigt werden, als dass immerhin 60 % der

Vpn die kombinierten Strategien bevorzugen. Allerdings ist auch die V-Strategie bei 40 % aller Vpn sehr beliebt. Die D-Strategie spielt hingegen keine Rolle, da sie von keiner einzigen Vp präferiert wird (siehe Tabelle 5.4).

Auswertung 2: Konsistenz der präferierten Strategie

Im Anschluss an Auswertung 1 haben wir die Daten jeder einzelnen Vp noch ein zweites Mal ausgewertet, und zwar hinsichtlich der Konsistenz in der Anwendung der von ihr (gemäss Auswertung 1) präferierten Suchstrategie. Diesmal überprüften wir jede der 70 Aufgaben hinsichtlich der Übereinstimmung in der Suchreihenfolge gemäss der Rangfolge der entsprechenden Gütekriterien der präferierten Suchstrategie. Die Häufigkeiten der eindeutigen und nicht-eindeutigen (andere Suchstrategien nicht ausschliessenden) Anwendungen sind in Abbildung 5.2 abgetragen. Das Kriterium einer konsistenten Anwendung setzten wir bei mindestens der Hälfte der Aufgaben, in denen überhaupt nach Informationen gesucht wurde. Von den 20 Vpn erfüllen lediglich 7 das definierte Konsistenz-Kriterium (siehe Abbildung 5.2). Wir können also nur für 35 % der Vpn statuieren, dass sie eine ganz bestimmte Informationssuchstrategie in der Mehrzahl der Aufgaben anwenden.

Von den 7 Vpn die das Präferenz-Kriterium erfüllen, verwendeten 4 vorwiegend die Success-Strategie und 3 die Validitäts-Strategie. Gemäss Auswertung 2 bleibt also nur noch ein leichter Vorteil für die kombinierten Strategien (siehe Hypothese 3) bestehen. Betrachten wir alle 20 Versuchspersonen, so ist das Verhältnis zwischen der Validitäts- und der Success-Strategie praktisch ausgeglichen (9 x Success- vs. 8 x Validitäts-Strategie). Nur 3 Vpn bevorzugen eine andere Suchstrategie als eine der beiden eben genannten: Die Vpn 7, 18 und 21 haben alle eine leichte Präferenz für die V*D-Strategie. Nichtsdestotrotz müssen wir konstatieren, dass die Strategie-Präferenz inter-personal deutlich variiert. Wir können ganz klar davon ausgehen, dass es keine von allen Vpn gleichermassen bevorzugte Informationssuch-Strategie gibt. Die beiden am stärksten präferierten Suchstrategien sind aber eindeutig die Validitäts- und die Success-Strategie. Interessant daran ist, dass es sich dabei um je eine einfache und eine kombinierte Suchstrategie handelt. Auch intra-personal finden wir eine grosse Varianz in der Suchstrategie-Anwendung: Die Vpn haben die von ihnen präferierte Suchstrategie in der Regel nicht sehr konsequent angewendet (siehe Tabelle 5.4, Phase 1).

Unseren Erwartungen hingegen entsprechend fallen die Daten bezüglich der Anwendung von One-Reason Decision Making (ORDM) aus: Alle 20 Vpn zusammen betrieben in insgesamt 46.5 % aller Aufgaben ORDM, d.h. sie stoppten ihre Informationssuche nach dem Aufdecken der ersten diskriminierenden Cue-Information. In rund 90 % dieser Fälle entschieden sie sich anschliessend erwartungsgemäss in Richtung der diskriminierenden Cue-Information. Damit unterstreicht auch dieses Experiment die Bedeutung von einfachen Entscheidungsregeln in Entscheidungsprozessen.

5.3.3 Diskussion

Die Ergebnisse dieses Experiments sprechen klar dagegen, dass es eine von allen Vpn gleichermassen bevorzugte Strategie gibt, an der sie sich bei der Informationssuche orientieren. Ent-

scheidungen werden in der Regel nicht so getroffen, wie wir in der starken Hypothese formuliert haben. Hingegen gibt es einige Vpn, die ein intra-individuell mehr oder weniger konstantes Suchverhalten aufweisen. Von den fünf theoretisch postulierten und in diesem Experiment untersuchten Suchstrategien stehen zwei ganz klar im Vordergrund: Die Validitäts- und die Success-Strategie. Die Success-Strategie erweist sich dabei als eine ernstzunehmende Konkurrenz zur Validitäts-Strategie, die als Suchstrategie, im Rahmen des One-Reason Decision-Makings, der TTB-Heuristik von Gigerenzer entspricht (vgl. z.B. Gigerenzer & Todd, 1999).

Die Konsistenz in der Anwendung dieser beiden Suchstrategien liegt allerdings unter unseren Erwartungen, da wir lediglich 7 der insgesamt 20 Vpn klar als Anwender einer dieser beiden Suchstrategien qualifizieren konnten, wobei das Verhältnis zwischen Success- und Validitäts-Strategie 4:3 beträgt und damit beinahe ausgeglichen ist. Auch das individuelle Informationssuch-Verhalten der anderen 13 Vpn ordneten wir jeweils einer der fünf postulierten Suchstrategien zu. Allerdings verwendete jede dieser 13 Vpn die bei ihr am häufigsten beobachtete Strategie in weniger als 50% der Aufgaben in denen sie überhaupt nach Informationen gesucht hatte. Deshalb können wir bei diesen Vpn jeweils nicht von einer konsistent angewendeten Suchstrategie sprechen.

5.4 Gesamtdiskussion

Die Verwendung einfacher Urteilsheuristiken im Sinne eines übergeordneten *One-Reason Decision Making* kann für beide Experimente bestätigt werden. Damit steht bei Entscheidungen unter Unsicherheit eine einfache Stopppregel im Vordergrund, welche den Umfang der Informationssuche bestimmt: Sobald eine diskriminierende Information vorhanden ist, kann definitiv entschieden werden. Die Richtung der Entscheidung (Entscheidungsregel) orientiert sich mit einer vernachlässigbaren Fehlerrate in ebenso einfachster Art gemäss dem Hinweis der vorliegenden Information. Soweit so gut!

Die Art der Suche (Suchrichtung bzw. Suchstrategie) scheint hingegen nicht so klar ausgeprägt zu sein. Während die Befundlage im ersten Experiment eher für eine multiplikative Verknüpfung der beiden Cue-Gütemasse Validität und Diskriminationsrate - also für *Usefulness* - spricht, überwiegt im zweiten Experiment der *Success*, und somit eine leichte Überbewertung der Validität. Und während im ersten Experiment die Validitätsstrategie (TTB) eine sehr geringe Rolle zu spielen scheint, ist sie im zweiten Experiment neben *Success* die am häufigsten angewendete Strategie.

Hier stellt sich die berechtigte Frage, ob die unterschiedlichen Entscheidungssituationen mit ein Erklärungsfaktor ist. Im Städteparadigma sind die Diskriminationsraten sinngemäss klein. Die Versuchspersonen könnten sich auf eine längere Suche nach einem diskriminierenden Hinweis eingestellt haben. Die normative Suchstrategie wäre entsprechend die gezeigte Orientierung nach der *Usefulness* der Cues ($V \cdot D$). Im Quiz-Setting sind die Diskriminationsraten dank der personalisierten Cues viel höher (vgl. Hausmann, 2004), womit die „Hoffnung“ auf einen schnellen diskriminierenden Hinweis mitspielen könnte. Diese Erwartung, dass im optimalen Fall bereits der

erstgewählte Cue diskriminiert (und dennoch eine relativ hohe Validität besitzt), wäre entsprechend mit der hierfür normativen *Success*-Orientierung der Cues zu rechtfertigen. Für Personen, denen es nicht darauf ankommt, mehrere „Cues abzugrasen“ und aber auf „Nummer sicher gehen“ wollen, wählen mit Vorteil eine Validitätsstrategie bzw. TTB. Diese Interpretation wäre ein Hinweis auf die flexible (adaptive) Anpassung der Suchstrategie an die entsprechende Umweltsituation.

Dennoch bleibt die Frage offen, weshalb die Versuchspersonen die präferierte Strategie nicht konsistenter anwenden, denn nur die Minderheit tut dies in mehr als 50% der Aufgaben mit Informationssuche. Ein erster berechtigter Einwand sind die geringen Unterschiede in den Gütemassen: Ist ein Cue mit einem Success von .65 ein viel besserer Cue als einer mit .62? Hier stellt sich die Frage einer möglichen Diskriminationsschwelle, also ob und wenn ja, ab welcher Differenz zwei Cues als identisch zu betrachten sind. Die Folge von identisch wahrgenommenen Cues wäre deren Vertauschbarkeit (vgl. die semiorde von Luce, 1956). Ein weiterer Einwand betrifft die vorhandenen mathematischen Verarbeitungsfehler bei der Integration der vorgelegten Gütemasse. Vielleicht sind Menschen gar nicht dazu in der Lage (wie u.a. von Kahneman & Tversky gezeigt), die beiden präsentierten Werte V und D hinsichtlich der postulierten Gütemasse *Success* und *Usefulness* korrekt zu verrechnen. Die Folge beim Eintreffen beider Einwände wäre eine falsche bzw. Nicht-Detektion von Suchstrategien in den vorliegenden Experimenten.

Die Suchregel von *Take the Best* (Validitätsstrategie) im strikten Sinne kann in Experiment 2 nur bedingt und in Experiment 1 (im Vergleich zu den übrigen Strategien) nicht gefunden werden: Es ist nicht primär die Validität, die den Kaufanreiz für eine Cue-Information bildet. In Experiment 1 wird die Diskriminationsrate (D) doppelt so häufig in Form einer möglichen Suchstrategie angewendet als die reine Validitätsstrategie. Der hohe Anteil an reiner D-Strategie könnte ein methodisches Artefakt sein, denn die Überlappung mit anderen Strategien ist bei D höher als bei V. In Experiment 2 wurde eine reine D-Strategie von keiner der Vpn als präferierte Strategie gewählt.

Insgesamt können Suchstrategien, welche auf einer Kombination von Validität und Diskriminationsrate aufbauen, mehr Fälle erklären als eine einseitige Ausrichtung auf V oder D. Die Ping-Pong-Strategie aus Experiment 1 ist (trotz hoher Fallzahlen in Experiment 1) keine ernst zu nehmende Suchstrategie, denn dieses Verfahren hat die meisten „Freiheitsgrade“: Vor allem wenn der erste gekaufte Cue diskriminiert, überlappt die Vorhersage in fast allen Fällen mit den anderen Strategien. Somit verbleiben die additive und die multiplikative Verknüpfung, sowie diejenige nach der Success-Formel. Einen Hinweis auf die Verwendung der additiven Verknüpfung hat der Härtetest (Experiment 2) in der untersuchten Stichprobe ebenfalls nicht zu tage gefördert. Die „Sieger-Strategien“ heißen also *Usefulness* und *Success*, wobei dem Situationsbezug unbedingt Rechnung getragen werden muss, d.h. in einer Umwelt mit beschränkter Informationssuche wäre eher *Success* und bei einer unbeschränkten eher *Usefulness* als Suchstrategie angebracht.

Im negativen Sinne aufschlussreich ist jedoch auf jeden Fall der Befund, dass sich keine Person auch nur in der Hälfte der Fälle des angewandten *One-Reason-Decision-Makings* an die Validität als Kaufargument für einen Cue hält. Will *Take-the-Best* also den Anspruch erheben, ein deskriptives Modell menschlichen Entscheidens unter den Bedingungen ökologischer Rationalität zu sein (sequentielle Informationssuche sowie Informationskosten), dann muss nach den hier vorgelegten Daten der Formalismus für die Festlegung der Güte der Cues erweitert werden. Sowohl die zahlreichen Simulationsstudien der ABC-Gruppe um Gigerenzer (Gigerenzer et al., 1999) als auch das von Bröder (2000a, 2000b) berichtete Validitätstraining setzen letztlich auf das falsche Pferd: Der „gute Cue“ ist in den Augen von Personen, die für Information zu zahlen haben (und das ist die natürliche Situation im Ansatz der Ökologischen Rationalität) in Experiment 1 derjenige Cue, der eine möglichst hohe Diskriminationsrate und eine möglichst hohe Validität aufweist. Der Cue mit der höchsten Validität ist lediglich zweiter Sieger und wird nur in Kombination mit hoher Diskriminationsrate gekauft oder dann, wenn der am höchsten diskriminierende Cue eben in diesem Fall nicht zur gewünschten Information führte und deswegen eine zweite Investition getätigt werden muss. Dieses Ergebnis ist vereinbar mit der Rahmentheorie der Ökologischen Rationalität: Wenn in der Umwelt Suchkosten entstehen, dann entpuppt sich die Suche nach validen und diskriminierenden Cues als eine adaptive Strategie.

Vernachlässigt man allerdings die Suchregel und betrachtet TTB nur als *One-Reason-Decision-Making* mit TTB-konformer Stoppregel und Entscheidungsregel, dann beschreibt eine solche Heuristik das Vorgehen der allermeisten Versuchspersonen recht gut.

Dieses Ergebnis wäre mit den Mitteln einer reinen Verhaltensvorhersage auf der Ebene der Entscheidungen übrigens nicht sichtbar gewesen: Würde man, so wie in allen Simulationsstudien und in den Auswertungen der bisherigen Experimente geschehen, auch diese Daten nach dem Zugang des *Structural Modeling* auswerten und dabei TTB als non-kompensatorische Cue-Gewichte hinsichtlich der Rangordnung der Validitäten verstehen, dann hätte *Take-the-Best* in Experiment 1 die stolze Bilanz von 89,4% korrekt vorhergesagter Urteile! Doch diese Zahl täuscht gewaltig, denn berücksichtigt man neben der Richtung der Entscheidung auch den Prozess der Informationssuche mit, dann sinkt die Vorhersagequote eines rein an der Validität orientierten TTB wie gesehen auf nurmehr 15%.

Deswegen möchten wir die Frage „How much do people pay for validity“ mit einem „nicht sonderlich viel“ beantworten. Und das scheint – aus der Sichtweise einer ökologischen Rationalität – auch seine guten Gründe zu haben: Personen, die bei Informationen eher in die hochgradig validen Cues investieren, haben dabei keinen sonderlichen Geschäftserfolg ($r = -.18$). Sie stehen – wegen Nichtberücksichtigung der Diskriminationsrate – nur allzuhäufig mit leeren Händen da, während diejenigen Personen, die gleich die Diskriminationsrate berücksichtigen, für weniger Geld Information bekommen. Zumindest in einem Setting mit hohen Informationskosten wiegt das die etwas häufigeren Fehlentscheidungen mehr als auf, wie eine positive Korrelation von $r = .74$ zwischen der Anzahl diskriminationsratenorientierter Cuekäufe und dem Geschäftserfolg zeigt. Wenn unsere Verhaltensstrategien tatsächlich so erfolgreich an die Umwelt angepasst sind wie das das Konzept der Ökologischen Rationalität verkündet, dann sollte auch die psychologi-

sche Theoriebildung den Nutzensaspekt der Diskriminationsrate von teurer Information stärker berücksichtigen. Menschen tun es jedenfalls.

5.5 Literatur

- Bröder, A. (2000a). „Take The Best – Ignore The Rest”: Wann entscheiden Menschen begrenzt rational? [“Take The Best - Ignore The Rest” When are people's decisions boundedly rational?]. Lengerich: Pabst Science Publishers.
- Bröder, A. (2000b). A methodological comment on behavioral decision research. *Psychologische Beiträge*, 42, 645–662.
- Bröder, A. (2000c). Assessing the empirical validity of the “Take The Best”-heuristic as a model of human probabilistic inference. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 26, 1332–1346.
- Bröder, A. (2001). Die eingeschränkte Fruchtbarkeit eines unvollständigen Forschungsprogramms: Kommentar zu Hertwig und Hoffrage [The restricted fruitfulness of an incomplete research program: Comment on Hertwig and Hoffrage]. *Psychologische Rundschau*, 52, 159–162.
- Bröder, A. (2002). Take The Best, Dawes' Rule, and compensatory decision strategies: A regression-based classification method. *Quality & Quantity*, 36, 219–238.
- Bröder, A. (2003). Decision making with the “adaptive toolbox”: Influence of environmental structure, intelligence, and working memory load. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 29, 611–625.
- Bröder, A. (2004). *Entscheiden mit der „Adaptiven Werkzeugkiste”: Ein empirisches Forschungsprogramm*. Lengerich: Pabst Science Publishers.
- Bröder, A., & Schiffer, S. (2003a). “Take The Best” versus simultaneous feature matching: Probabilistic inferences from memory and effects of representation format. *Journal of Experimental Psychology: General*, 132, 277–293.
- Bröder, A., & Schiffer, S. (2003b). Bayesian strategy assessment in multi-attribute decision research. *Journal of Behavioral Decision Making*, 16, 193–213.
- Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, 62, 193–217.
- Chater, N. (2000). How smart can simple heuristics be? Commentary on: simple heuristics that make us smart. *Behavioral and Brain Sciences*, 23, 745–746.
- Conolly, T., & Gilani, N. (1982). Information search in judgment tasks: A regression model and some preliminary findings. *Organizational Behavior and Human Performance*, 30, 330–350.
- Cooper, R. (2000). Simple heuristics could make us smart: but which heuristic do we apply when? Commentary on: simple heuristics that make us smart. *Behavioral and Brain Sciences*, 23, 746.
- Czerlinski, J., Gigerenzer, G., & Goldstein, D. G. (1999). How good are simple heuristics? In G. Gigerenzer, P. M. Todd & the ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 97–118). New York: Oxford University Press.
- Erdbauer, M. (Hrsg.) (2001). *Wissensportal Allgemeinbildung: Das aktuelle Wissen unserer Zeit in Frage und Antwort*. Trautwein Lexikon-Edition. München: Compact.
- Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, 103, 650–669.

- Gigerenzer, G., & Goldstein, D. G. (1999). Betting on one good reason: The Take The Best heuristic. In G. Gigerenzer, P. M. Todd & the ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 75–95). New York: Oxford University Press.
- Gigerenzer, G., & Todd, P. M. (1999). Fast and frugal heuristics: The adaptive toolbox. In G. Gigerenzer, P. M. Todd & the ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 3–34). New York: Oxford University Press.
- Gigerenzer, G., Czerlinski, J., & Martignon, L. (1999). How good are fast and frugal heuristics? In J. Shanteau, B. Mellers, & D. Schum (Eds.), *Decision Science and Technology: Reflections on the Contributions of Ward Edwards* (pp. 81–103). Norwell: Kluwer Academic Publishers.
- Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic Mental Models: A Brunswikan theory of confidence. *Psychological Review*, 98, 506–528.
- Gigerenzer, G., Todd, P. M., & the ABC Research Group (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.
- Hausmann, D. (2004). Informationssuche im Entscheidungsprozess. *Die Nützlichkeit von Hinweis-Cues und der Anspruch an Urteilssicherheit*. Dissertationsschrift. Zürich: Zentralstelle der Studentendruckerei.
- Hertwig, R., & Hoffrage, U. (2001a). Eingeschränkte und ökologische Rationalität: Ein Forschungsprogramm (Bounded and ecological rationality: A research program). *Psychologische Rundschau*, 52, 11–19.
- Hertwig, R., & Hoffrage, U. (2001b). Empirische Evidenz für einfache Heuristiken: Eine Antwort auf Bröder [Empirical evidence for simple heuristics: A response to Bröder]. *Psychologische Rundschau*, 52, 162–165.
- Hoffrage, U., Hertwig, R., & Czienskowski, U. (2003). *The ecological rationality of Take The Best's building blocks*. Vortrag an der 19th bi-annual conference on Subjective Probability Utility and Decision Making (SPUDM). Zurich: Swiss Federal Institute of Technology.
- Luce, R. D. (1956). Semiorders and a theory of utility discrimination. *Econometrica*, 24, 178–191.
- Martignon, L., & Hoffrage, U. (1999). Why does one-reason decision making work? A case study in ecological rationality. In G. Gigerenzer, P. M. Todd & the ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 119–140). New York: Oxford University Press.
- Martignon, L., & Hoffrage, U. (2002). Fast, frugal, and fit: Simple heuristics for paired comparison. *Theory & Decision*, 52, 29–71.
- Newell, B. R., & Shanks, D. R. (2003). Take the best or look at the rest? Factors influencing "one-reason" decision-making. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 29, 53–65.
- Newell, B. R., Rakow, T., Weston, N. J., & Shanks, D. R. (2004). Search strategies in decision-making: the success of "success". *Journal of Behavioral Decision Making*, 17, 117–137.
- Newell, B. R., Weston, N. J., & Shanks, D. R. (2003). Empirical tests of a fast and frugal heuristic: not everyone "takes-the-best". *Organizational Behavior and Human Decision Processes*, 91, 82–96.
- Oaksford, M. (2000). Speed, frugality, and the empirical basis of Take-The-Best. Commentary on: simple heuristics that make us smart. *Behavioral and Brain Sciences*, 23, 760–761.
- Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. Cambridge, MA: Cambridge University Press.
- Rieskamp, J., & Hoffrage, U. (1999). When do people use simple heuristics and how can we tell? In G. Gigerenzer, P. M. Todd & the ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 141–167). New York: Oxford University Press.
- Saad, G., & Russo, J. E. (1996). Stopping criteria in sequential choice. *Organizational Behavior and Human Decision Processes*, 67, 258–270.

-
- Sargent, T. J. (1993). *Bounded Rationality in Macroeconomics*. New York: Oxford University Press.
- Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, 69, 99–118.
- Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, 63, 129–138.
- Simon, H. A. (1982). *Models of bounded rationality*. Cambridge, MA: MIT Press.
- Todd, P. M., & Gigerenzer, G. (2000). Précis of simple heuristics that make us smart. *Behavioral and brain sciences*, 23, 727–741.
- Tversky, A., & Kahneman, D. (1974). *Judgment under uncertainty: Heuristics and biases*. *Science*, 185, 1124–1131.

Kapitel 6

Und sie lernen es doch! Das Einschätzen der Validität von probabilistischen Hinweis-Cues und der Einfluss der subjektiven Repräsentation auf das Urteilsverhalten

Das Konzept der Cue-Validität beim probabilistischen Entscheiden setzt als psychologisches Modell für menschliches Verhalten voraus, dass der Zusammenhang zwischen Prädiktor- und Zielvariable von Menschen aufgrund wiederholter Beobachtung frequentistisch korrekt eingeschätzt werden kann. Das Lernen von Cue-Validitäten wurde in einem Quiz-Szenario umgesetzt, in dem die Vpn beim Bearbeiten von schwierigen Aufgaben gleichzeitig beobachten konnten, wie gut andere Personen (Telefonjoker) die richtige Lösung trafen (Phase I). Es zeigte sich, dass die Cue-Validität (Expertise der Joker) ordinal richtig eingeschätzt wurde. Gleichzeitig wurden jedoch die absoluten Validitäten systematisch und massiv unterschätzt. In der anschließenden Anwendungsphase (II) waren die Vpn beim Beantworten der extrem schwierigen Quiz-Fragen auf sich alleine gestellt, konnten aber so viele Telefonjoker ihrer Wahl um ihren Tipp befragen. Hier erwies sich One-Reason Decision-Making als eine Funktion der eingeschätzten Validität: Wer ein besonders hohes Zutrauen in die Expertise des subjektiv validesten Telefonjokers gewonnen hatte, befragte wiederholt diesen alleine nach seinem Tipp, während sich Vpn mit einer niedrigeren Validitätseinschätzung systematisch auf die Tipps mehrerer Telefonjoker abstützten.

6.1 Einleitung

Die Lernfähigkeit des Menschen ist unbestritten. Ebenso unbestritten ist die Tatsache, dass wir in konkreten Entscheidungssituationen unsere bisherigen Erfahrungen und das Vorwissen über mögliche Alternativen und deren Merkmalsdimensionen einfließen lassen mit dem Ziel, eine möglichst gute Wahl zu treffen. Da wir im Alltag aber nicht im Voraus wissen können, welche Entscheidungssituationen auf uns zukommen werden, stellt sich die grundsätzliche Frage, wie viel wir aus der täglichen Beobachtung unserer Umwelt lernen können und wie wir das Gelernte in einer Entscheidungssituation konkret umsetzen.

Im Zusammenhang mit adaptivem Entscheidungsverhalten unter Unsicherheit besteht mittlerweile eine ganze „Toolbox“ von einfachen Urteilsheuristiken (u. a. Gigerenzer & Selten, 2001). Die theoretischen Implikationen auf den individuellen Lern- und Entscheidungsprozess sind allerdings empirisch erst ansatzweise überprüft worden (vgl. Hausmann & Läge, 2005a). Bislang wenig untersucht ist das komplexe Zusammenwirken von Vorwissen und subjektiven Erfahrungen auf die einzelnen Phasen des Entscheidungsprozesses, d.h. der Art und des Umfangs der Suche nach Informationen, des Zeitpunktes des Abbruchs der Informationssuche und der Art der Verarbeitung der gesuchten Informationen, auf individuellem Versuchspersonenniveau.

Eine ausführliche Theorie Probabilistischer Mentaler Modelle (PMM) haben Gigerenzer, Hoffrage & Kleinböting (1991) formuliert. Ihr kognitives Modell und die daraus abgeleiteten Urteilsheuristiken greifen im Wesentlichen auf die Konzepte von Herbert A. Simon und Egon Brunswik zurück. Simon (1955, 1956) hat darauf hingewiesen, dass die menschliche Rationalität beschränkt ist. Begrenzte menschliche Ressourcen wie Suchkosten, Zeitlimiten und beschränkte kognitive Verarbeitungskapazität führen zu einem heuristischen Entscheidungsverhalten. Es geht nicht darum, die eine optimale Option zu finden, sondern es reicht aus, eine hinreichend gute Wahl zu treffen. Brunswik (1952) beschreibt die indirekte Beziehung zwischen dem ökologischen System der Umwelt und des kognitiven Systems des Menschen unter Berücksichtigung der Unsicherheit in einem *Linsen-Modell*. Theoretisch betrachtet können wir Objekte, Ereignisse und Zustände in unserer Umwelt nie direkt erfahren. Wir konstruieren vielmehr eine mentale Repräsentation darüber, was dann die Grundlage für ein schlussfolgerndes Urteilen darstellt. Menschen beobachten Variablen, die einen Hinweis auf die wahre Ausprägung geben können, wie beispielsweise Regenwolken oder tieffliegende Schwalben, welche auf das Ereignis „baldiger Regen“ hindeuten. Diese sogenannten Hinweisreize (sogenannte Cues) stehen in einer mehr oder weniger engen Beziehung zum wahren bzw. distalen Zustand. Sie sind also unterschiedlich „ökologisch valide“. Statistisch betrachtet kann die Enge der Beziehung zwischen jedem proximalen Hinweisreiz und der distalen Variable als Korrelation ausgedrückt werden. Die „wahre“ ökologische Validität kennt nur derjenige – wie z. B. der erfahrene Meteorologe –, der die „wahren“ Beziehungen zwischen proximaler und distaler Variable kennt. Ein naiver Beobachter wird die Korrelation subjektiv einschätzen, je nach persönlicher Erfahrung oder momentanem Eindruck. Es ist aber davon auszugehen, dass sich die subjektiven Cue-Validitäten mit zunehmender Erfahrung den ökologischen annähern.

Das Verdienst von Gigerenzer et. al. (1991) ist es, die Ansätze von Simon und Brunswik reformuliert und in einen übergeordneten Theorierahmen des adaptiven Urteilsverhaltens unter Unsicherheit gestellt zu haben. Gigerenzer und die ABC Research Group haben ausgehend von ihrer PMM-Theorie in den vergangenen 10 Jahren diverse Urteilsheuristiken aufgestellt (vgl. Übersicht bei Todd & Gigerenzer in Gigerenzer, Todd & the ABC Research Group, 1999) von denen der prominenteste Algorithmus als Take The Best-Heuristik (TTB) bezeichnet³¹ wurde.

Gemäss TTB (Gigerenzer & Goldstein, 1996) werden Probability-Cues entsprechend ihrer ökologischen Validitäten in absteigender Höhe hierarchisch geordnet und bei der Informationssuche (intern oder extern) sequentiell nach den binären Ausprägungen für die zur Auswahl stehenden Alternativen abgefragt. Falls eine Aussage differenziert im Sinne von „die eine Alternative hat die Ausprägung und die andere nicht“, liegt die Grundlage für ein Urteil vor. Diskriminiert der valideste Cue nicht, wird der zweitvalideste Cue befragt etc. Die TTB-Heuristik ist non-kompensatorisch und die Abfrage folgt einer lexikographischen Ordnung. TTB gehört in die Gruppe der Urteilsheuristiken, in der Entscheidungen anhand einer einzigen Information bzw. eines einzigen diskriminierenden Cues getroffen werden. Eine Entscheidung nach TTB wird deshalb auch als ein „One-Reason Decision Making“ (ORDM) bezeichnet (Gigerenzer & Goldstein, 1999, S. 79).

Die Gruppe um Gigerenzer hat in verschiedenen Simulationen nachgewiesen, dass TTB eine ebenso gute Performance aufweist wie andere klassische Entscheidungsmodelle (Czerlinski, Gigerenzer & Goldstein, 1999; Gigerenzer & Goldstein, 1996; 1999; Martignon & Hoffrage, 1999). TTB trifft spezifische Vorhersagen über die einzelnen Phasen des Entscheidungsprozesses (vgl. Argumentation über den Prozess der Entscheidung in Hausmann, 2004) und deren Anwendung kann daher im Bezug auf die einzelnen building blocks (Such-, Stopp- und Entscheidungsregel) empirisch überprüft werden (vgl. auch Kritik von Bröder, 2000a; 2000c; 2004). Empirische Untersuchungen von Bröder (2000a; 2000c; 2002; 2003; 2004), Bröder & Schiffer (2003), Newell & Shanks (2003); Newell, Weston & Shanks (2003); Newell, Rakow, Weston & Shanks (2003), sowie Hausmann (2004), (vgl. auch Chater, Oaksford, Nakisa & Redington, 2003), haben denn auch mit unterschiedlichen Methodenansätzen gezeigt, dass nicht alle Versuchspersonen TTB anwenden. In Läge, Daub, Christen & Hausmann (2005) und Hausmann & Läge (2005b) wurden vor allem Verstösse gegen die Suchregel von TTB festgestellt, da ein Teil der Versuchspersonen die Höhe der Diskriminationsrate mit der Validität eines Cues verrechnen und daraus individuell eine differenziertere Suchstrategie bilden (vgl. Hausmann & Läge, 2005a). Verstösse gegen die Stoppregel haben praktisch sämtliche oben erwähnten empirischen Studien festgestellt (vgl. Hausmann & Läge, 2005c). Obwohl die einzelnen Untersuchungen zur empirischen Überprüfung von TTB weder methodisch noch inhaltlich identisch strukturiert sind, ist dennoch als Zusammenfassung der Befunde festzuhalten, dass „Take The Best“ keine universell angewendete Heuristik ist (vgl. auch Chater et. al., 2003). Die Suche nach einschränkenden

³¹ TTB wurde zuerst als Take-The-Best-Algorithmus Gigerenzer & Goldstein (1996) und später als Take-The-Best-Heuristik Gigerenzer et. al. (1999) bezeichnet.

bzw. TTB begünstigenden Faktoren hat denn bereits einige wichtige Erkenntnisse geliefert. TTB - wie generell auch One-Reason Decision Making - ist abhängig von der Höhe der Suchkosten im Vergleich zum möglichen Gewinn (vgl. Hausmann, Christen & Läge, 2005); Übersicht der entsprechenden Studien in Hausmann & Läge, 2005a. Eine ausführliche Aufstellung von weiteren situativen Faktoren gibt Hausmann, 2004.

Die zentrale Voraussetzung des Cue-Erwerbs für die Anwendung von TTB wurde bislang empirisch nur ungenügend berücksichtigt. Gigerenzer et. al. (1991) gehen davon aus, dass die Abfrage von Cues gedächtnisbasiert verläuft. Die PMM-Theorie macht explizite Aussagen über die Repräsentation von Cue-Validitäten: „(...) cue validities are learned by observing the frequencies of co-occurrences in an environment“ (p. 510). Damit gibt Gigerenzer und seine Kollegen eine Erklärung dafür, wie subjektive Konfidenzen inzidentell gelernt werden können. Das Lernen von ökologischen Cue-Validitäten beinhaltet demnach, die relative Häufigkeit korrekter Urteile anhand eines Cues intern zu repräsentieren. Die Validität eines binären Cues ist die relative Häufigkeit korrekter Urteile anhand dieses Cues, wenn er zwischen den Alternativen diskriminiert. Verschiedene Untersuchungen haben gezeigt, dass Menschen fähig sind, solche relativen Häufigkeiten ziemlich akkurat einzuschätzen³² (Attneave, 1953; Shuford, 1961; Peterson & Beach, 1967; Estes, 1976; Hasher & Zacks, 1979; 1984; Bröder, 2000a; 2000c; Newell, 2000). Gegen eine automatische Registrierung aller denkbaren Stimulusklassen argumentieren vor allem evolutionspsychologisch orientierte Forscher (Brase, Cosmides & Tooby, 1998; Martignon & Hoffrage, 1999). Im Zusammenhang mit der Überprüfung von TTB fand allerdings bislang kein konsequentes Cue-Lernen statt!

Beim experimentellen Nachweis von TTB wurden die Cue-Validitäten in einer vorgängigen Lernphase anhand einer Serie von Einzelentscheidungen - in einem sogenannten Cue-Training - vermittelt. Beim Lösen von einzelnen Entscheidungsaufgaben mit zwei oder mehr Alternativen werden die Versuchspersonen aufgefordert, das Antwortverhalten der Cues mitzuverfolgen. Entsprechend ihrer ökologischen Validitäten werden die Cues unterschiedlich viele korrekte Vorhersagen treffen, was wiederum von der Versuchsperson frequentistisch beobachtet werden kann (bei Ausbalancieren der diskriminierenden Cues über alle Cueantworten). Bröder (2000a; 2000c) hat einzelne Cues in Blöcken lernen lassen (was bei fünf Cues und 50 Lerndurchgängen pro Cue bis zu 250 Einzeldurchgänge erforderte), während z.B. Newell, Weston & Shanks (2003) jeweils alle Cues simultan präsentierte und somit lediglich 50 Einzeldurchgänge benötigte. Die Versuchsperson wird am Schluss der Lernphase aufgefordert, die Güte der Cues in einem Rating zueinander in Beziehung zu setzen. Es wird einerseits davon ausgegangen, dass diese „gelernte“ Rangfolge der Cue-Validitäten der intern repräsentierten entspricht, welche dann das Suchverhalten in der Anwendungs- bzw. Testphase steuert und auf die Verwendung von TTB hin überprüft werden kann. Andererseits wird davon ausgegangen, dass die frequentistischen Beobachtungswerte in Vorhersagewahrscheinlichkeiten (subjektive Validitäten) umwandelt werden. In

³² Gemäss Estes (1976) sind subjektive Einschätzungen nur dann verzerrt, wenn Cues in der Lernphase nicht gleich häufig auftreten.

dieser Testphase können die Cues aktiv und sequentiell um ihre aktuelle Ausprägung abgefragt werden, die Reihenfolge und der Umfang der Abfrage ist dabei der Versuchsperson freigestellt.

Aus Gründen der Vergleichbarkeit in der Testphase war man allerdings darauf angewiesen, dass die ökologischen Validitäten korrekt gelernt wurden. Um das aufwendige Cue-Lernen abzukürzen, sind Bröder und Newell in ihren Untersuchungen dazu übergegangen, den Versuchspersonen im Verlauf der Lernphase einen Hinweis (Hint) zu geben. Bei Bröder (2000a; 2000c) war dies jeweils ein visuelles Feedback in Form eines graphischen Balkens für jeden einzelnen Cue, der die relative Häufigkeit der korrekten vs. der falschen Antworten abbildete. Newell & Shanks (2003) und Newell, Weston & Shanks (2003) teilten den Versuchspersonen die korrekte Rangordnung der Cue-Validitäten in der Mitte und zum Schluss der Lernphase mit. Es ist deshalb nicht erstaunlich, dass in der anschließenden Testphase praktisch sämtliche Versuchspersonen die korrekte Suchregel von TTB (nicht aber die Stopppregel) verwendeten. In den Untersuchungen zu den Suchstrategien und Suchkosten in Hausmann (2004) wurde gänzlich auf die Lernphase verzichtet und den Versuchspersonen die wahren ökologischen Werte von Validität und Diskriminationsrate direkt und jederzeit sichtbar präsentiert.

Bisherige Untersuchung von Bröder (2000a; 2000c), Newell & Shanks (2003) und Newell, Weston & Shanks (2003) untersuchten das Entscheidungsverhalten auf Stichprobenebene, ohne die individuellen Lernunterschiede mit zu berücksichtigen. Will man TTB im Rahmen der PMM-Theorie überprüfen, dann ist es eine berechtigte Forderung, die Subjektivität der Versuchspersonen im Bezug auf die Lern- und Entscheidungsvariablen konsequent mit zu berücksichtigen. Wichtig ist es, die aufgrund des Häufigkeits-Lernens aufgebaute Cue-Repräsentation einer Person als individuelle Basis für das Entscheidungsverhalten zu betrachten. Konsequenterweise muss TTB hinsichtlich der *subjektiven* Repräsentationen von Cue-Rangfolgen und Cue-Validitäten auf *individuellem* Versuchspersonenniveau getestet werden. Der Einfluss von repräsentierter Cue-Güte auf das Entscheidungsverhalten kann nur dadurch fair festgestellt werden, wenn Such-, Stopp- und Entscheidungsregel individuell, d.h. hinsichtlich der subjektiv repräsentierten Cue-Validitäten und Cue-Rangfolge getestet werden. Dies verlangt eine individuelle Auswertung der Verhaltensmuster. Im Bezug auf die Anwendung von TTB in einem alltagsnäheren Setting ist es zudem notwendig, das individuelle Vorwissen bezüglich der Cues und der Alternativen zu kontrollieren bzw. mit zu messen. Für die vorliegende Untersuchung wählten wir ein Setting, bei dem kein Vorwissen bezüglich der Cues existierte, bezüglich der Alternativen hingegen schon. In einer Urteilssituation muss die Versuchsperson jeweils die korrekte Antwortalternative zu einer realen Quizfrage wählen, wobei das jeweilige Vorwissen durch Konfidenzurteile abgefragt wurde (vgl. Hausmann, 2004). Die Cues sollten als anonymisierte Informanten dargestellt werden, deren Expertenwissen bezüglich Quizfragen vor der Trainingsphase völlig unbekannt sind und deren Expertise (Validität) erst im Laufe mehrerer Durchgänge gelernt werden müssen.

Da die Cue-Validitäten bisher im Laufe der Lernphase „menu-based“ vermittelt wurden, d.h. das angebliche „Cue-Lernen“ durch visuelles Feedback Bröder (2000a; 2000c) oder durch explizite Hinweise (Hints) Newell & Shanks (2003) vereinfacht wurde, fehlt die Überprüfung von zwei

wesentlichen Voraussetzungen für TTB: Erstens die adäquate subjektive Repräsentation von ökologischen Cue-Validitäten anhand einer reinen Häufigkeits-Erfahrung und zweitens die korrekte Umsetzung der subjektiv repräsentierten Cue-Validitäten in einer konkreten Entscheidungssituation im eigentlichen Sinne von „take the best (Favorisierte Suche des Cues mit der höchsten Cue-Validität)“. Mit einem einfachen „memory-based“ inference Design wollen wir die Güte der gelernten Cues pro Versuchsperson bestimmen und das Entscheidungsverhalten konsequent auf die subjektiv gelernten Cues und deren Validitäten hin auswerten. Mit möglichst optimalen situativen Parametern (Höhe der Suchkosten, Vermeiden von nicht diskriminierender Information, Kontrolle des Vorwissens, natürliches Setting) soll sichergestellt werden, dass die Anwendung von TTB allein aufgrund der subjektiv gelernten Cue-Validitäten betrachtet werden kann. Wir verfolgen zwei Fragestellungen:

Fragestellung 1: Wie gut lernen Personen Cues und ihre Validitäten? (Aufbau einer Repräsentation)

Das Anwenden von TTB verlangt eine klare, ordinal repräsentierte Rangfolge von Cues. Wir wollen überprüfen, ob diese Rangfolge von Cues innerhalb kurzer Lerndurchgänge aufgrund von rein frequentistischer Beobachtung der korrekten und falschen Antworten der Cues gebildet werden kann. Wir vermuten, dass die relationale Beziehung der Cues zueinander besser gelernt wird als die quantitativen Ausprägungen der einzelnen Cue-Validitäten (V). Zudem gehen wir davon aus, dass objektiv nahe beieinanderliegende Cue-Validitäten aufgrund von Interferenz schlechter gelernt werden. Die Abstände der ökologischen Cue-Validitäten (V_{diff}) sollen deshalb variiert werden. Um ein optimales Cue-Lernen zu gewährleisten, beschränken wir uns auf ein reines Validitäts-Training und setzen die Diskriminationsrate (D) auf ihre maximale Ausprägung ($D = 1$).

Fragestellung 2: Gibt es einen Einfluss der wahrgenommenen Cue-Güte auf die Anwendung von TTB? (Nutzung der Repräsentation im Sinne von TTB)

Die Höhe der subjektiv repräsentierten Cue-Validitäten und die Beziehung zueinander bestimmen, ob TTB in einer konkreten Entscheidungssituation von den Versuchspersonen angewendet wird oder nicht. Gemäss der Suchregel von TTB wird der subjektiv valideste Cue zuerst abgerufen etc. Wenn Cue-Validitäten als ähnlich oder gleich valide eingeschätzt werden, besteht die Gefahr der Verletzung der Suchregel, indem Cues mit ähnlich hohen Cue-Validitäten in einer konkreten Entscheidungssituation vertauschbar werden (vgl. Luce, 1956).

Gemäss der Stoppregel von TTB wird die Suche nach weiteren Informationen dann abgebrochen, sobald ein Cue zwischen den Alternativen diskriminiert. Im Idealfall also schon nach dem ersten Cue. Dabei spielt es gemäss TTB keine Rolle, wie hoch die Ausprägung dieses validesten Cues ist. Alternative Konzepte wie die der gewünschten Urteilssicherheit von Chaiken, Liberman & Eagly (1989) legen nahe, dass die Stoppregel bei Cues mit niedrig repräsentierten Cue-Validitäten unterhalb eines individuellen Niveaus missachtet werden könnte (vgl. Hausmann & Läge, 2005c). Falls der valideste Cue das gewünschte Mass an Urteilssicherheit nicht erreicht,

wird nach zusätzlicher Information gesucht und somit nicht mehr non-kompensatorisch entschieden.

6.2 Methoden

32 Studierende der Philosophischen Fakultät der Universität Zürich und vereinzelte Personen aus dem Berufsleben (31% female, mean age = 26.6, range = 20 bis 34) lösten insgesamt 70 schwierige Quizfragen mit je 4 Antwortalternativen. Die Computeroberfläche (vgl. Abbildung 6.1) (Visual Basic, 6.0) wurde von einem früheren Experiment übernommen (vgl. Läge, Daub, Christen & Hausmann, 2005) und entspricht in den Grundzügen dem mouselab von Payne, Bettman & Johnson (1993). Die Versuchspersonen hatten die Aufgabe, möglichst viele der kniffligen Wissensfragen korrekt zu beantworten. Für jede korrekt beantwortete Aufgabe erhöhte sich der Spielstand um 1'000 Punkte.

The screenshot displays a quiz application interface with four main panels:

- Frage (Question):**

Wem diente Kardinal Mazarin als leitender Minister?

Options:

 - Ludwig XVI. (A)
 - Karl der Kühne (B)
 - Napoleon (C)
 - Ludwig XIV. (D)

Wie sicher bist Du? 26% - 60% (ziemlich unsicher) [OK]
- Kontostand - Validitätstraining (Score Panel):**

Einnahme bei richtiger Antwort: 1'000
 Bisheriger Kontostand: 3'000
 Ausgaben bei aktueller Frage: 0
 Einnahme bei aktueller Frage: 0
 Resultierender Gewinn / Verlust: 0
 Neuer Kontostand: 3'000
 Antworten Total: 3 (falsche: 1 / richtige: 2)
- Telefonjoker (Joker Table):**

Joker	Antwort ist ...
A	richtig
B	falsch
C	richtig
D	richtig

[Weiter]
- Antwort (Erläuterung) (Answer Explanation):**

Deine Antwort ist leider falsch!

Kardinal Mazarin diente dem minderjährigen Ludwig XIV. (1638-1715), der erst 1660 die Macht übernahm, als leitender Minister. Mazarin förderte den Katholizismus in Frankreich. Auf ihn folgte Richelieu.

Abbildung 6.1: Eine Entscheidungsaufgabe in der Lernphase. Das Antwortverhalten der vier Joker (A-D) wird erst sichtbar, nachdem die Versuchsperson die Frage bereits ein erstes Mal beantwortet hat.

Die Quizfragen wurden aus unterschiedlichen Wissensgebieten zusammengestellt (Erdbauer, 2001; Strerath-Bolz, 2001) und vorgängig auf ihren Schwierigkeitsgrad hin getestet (vgl. Läge et. al., 2005). Für das Experiment wurden nur besonders schwierige Quizfragen ausgewählt. Nach der Präsentation einer Quizfrage, musste die Versuchsperson angeben, welche der vier

vorgegebenen Antwortalternativen sie präferierte. Die subjektive Urteilssicherheit dieser Präferenz musste durch die Angabe eines der fünf vorgegebenen Konfidenzintervalle eingeschätzt werden: 25% (völlig geraten), 26%-60% (ziemlich unsicher), 61%-90% (eher sicher), 91%-99% (ziemlich sicher) und 100% (absolut sicher). Damit wurde implizit das Vorwissen der Versuchsperson hinsichtlich der aktuellen Quizfrage getestet.

Die Probability-Cues wurden als vier voneinander unabhängige Personen mit unterschiedlichem Expertisegrad präsentiert. Diese „anonymisierten“ Experten, die in der 2. Phase auch als „Telefonjoker“ eingesetzt werden konnten, unterschieden sich darin, wie viele Quizfragen sie unabhängig vom Wissensgebiet korrekt beantworten konnten. Die Merkmalsausprägung eines Personen-Cues entsprach dem Tipp hinsichtlich einer der vier Antwortalternativen. Dieser Tipp wurde in der Lernphase lediglich in binärer Form angegeben: Ein Experte hatte die jeweilige Quizfrage entweder „richtig“ oder „falsch“ beantwortet. Die ökologischen Cue-Validitäten dieser personifizierten Cues entsprachen den Häufigkeiten korrekter Antworten über sämtliche Quizfragen und Wissensgebiete hinweg gesehen und entsprechend der mittleren Wahrscheinlichkeit korrekter Vorhersagen. Die Diskriminationsrate betrug 100%, d.h. jeder der Personen-Cues gab zu jeder Quizfrage einen Antwort-Tipp ab (Hinweis auf eine der Antwortalternativen a, b, c oder d). Durch das simultane Beobachten des Antwortverhaltens sämtlicher vier Personen-Cues über mehrere Durchgänge hinweg konnte sich die Versuchsperson eine subjektive Repräsentation über die Cue-Güte (Validität) aufbauen.

Mit der Quizshow (und entsprechend realen Quizfragen) als ökologische Umwelt sowie der (fiktiven) Telefonjoker als Cues wählten wir ein Entscheidungssetting, welches sich für die Versuchspersonen als besonders attraktiv herausstellte. Das individuelle Vorwissen zu jeder Quizfrage wurde mit kontrolliert, indem jeweils ein Konfidenzintervall angegeben werden musste. Der Vorteil personifizierter Cues für die Versuchspersonen besteht darin, dass das Konzept der Validität intuitiv als Trefferquote oder Zuverlässigkeit und die Diskriminationsrate als Anwendbarkeit von Cues vermittelt und leicht verstanden werden kann (vgl. Hausmann, 2004). Ein Cue in personifizierter Form entspricht zudem unserem sozialen Habitat, da wir in der Realität wichtige Hinweise bzw. Wissen meist mündlich oder schriftlich von anderen Personen vermittelt bekommen und diese Aussagen hinsichtlich ihrer Expertise und ihrer Anwendbarkeit überprüfen. Der Vorteil von personifizierten Cues für die experimentelle Manipulation liegt darin, dass sich die Höhe der Cue-Validitäten beliebig hoch und somit flexibel gestalten lässt. Merkmalsbasierte Cues in realen Entscheidungssettings sind selten unabhängig voneinander und können weder in der ökologischen Validität noch bei der Diskriminationsrate frei variiert werden. So sind die Cue-Validitäten wie z.B. beim Städtebeispiel von Gigerenzer & Goldstein (1996) durch die Objektverteilung vorgegeben. Die Gebundenheit an die Verteilung der Cue-Ausprägungen und Anzahl Objekte in der Referenzklasse gilt ebenso für die Diskriminationsrate, die bei vier Antwortalternativen maximal .67 betragen kann (p. 655). Bei personifizierten und voneinander unabhängigen Cues erhält die Diskriminationsrate eine etwas andere Bedeutung. Die Diskriminationsrate entspricht nun der Häufigkeit bzw. der Wahrscheinlichkeit, mit der ich überhaupt eine relevante Antwort auf die Frage bekomme. Liefert der Personen-Cue eine Antwort (er ist erreichbar), dann liegt eine diskriminierende Information vor, indem er auf eine der Antwortalternativen hinweist.

Gibt er aus irgendwelchen Gründen keine Antwort (er ist nicht erreichbar), liegt entsprechend kein brauchbarer Hinweis bzw. keine diskriminierende Information vor.

Die Versuchspersonen durchliefen zuerst eine Trainings- und anschliessend eine Anwendungsphase. Die Trainingsphase diente dazu, das Antwortverhalten von vier Telefonjokern zu beobachten und deren Expertisegrad (Validität) einzuschätzen, während in der Anwendungsphase die Joker dann auch entsprechend als Hilfestellung für die externe Informationssuche aktiv eingesetzt werden konnten.

Die ersten 50 Aufgaben dienten als Trainingsphase, bei der die Versuchsperson die jeweiligen Wissensfragen ohne fremde Hilfe beantworten musste. Nach der Präferenz- und Konfidenzintervallangabe gab der Computer eine Rückmeldung, ob die Frage richtig oder falsch beantwortet wurde und es erschien eine kurze Erläuterung zur entsprechenden Frage. Gleichzeitig wurde das Antwortverhalten der Joker simultan eingeblendet, also ob ein Joker die Frage richtig oder falsch beantwortet hat (vgl. Abbildung 6.1).

Die Wahl der ökologischen Validitäts-Werte erfolgte in Anlehnung an die Experimente von Bröder (2000a, S. 98). Als höchster Wert wurde $v_1 = 0.92$ und als tiefster Wert $v_4 = 0.52$ gewählt. Bei vier Antwortalternativen beträgt die Basiswahrscheinlichkeit bzw. die Zufallsrate .25. Die beiden mittleren Validitäts-Werte wurden in zwei verschiedenen Bedingungen variiert: Bedingung A ($v_2 = 0.76$, $v_3 = 0.60$) und Bedingung B ($v_2 = 0.84$, $v_3 = 0.68$). Die Bedingung A wurde von uns als leichter eingestuft, wenn es darum geht, den validesten Cue zu repräsentieren, denn der valideste Cue unterscheidet sich deutlicher vom 2. validesten. Entsprechend ähnliche Cue-Validitäten weisen die beiden höchsten Cues in Bedingung B auf. Die Versuchspersonen wurden zufällig einer der beiden Bedingungen zugeordnet. Die vier Joker wurden mit den Grossbuchstaben „A“, „B“, „C“ und „D“ bezeichnet und auf dem Bildschirm in dieser alphabetischer Reihenfolge präsentiert. Die Zuordnung der ökologischen Cue-Validitäten zu den Jokernamen erfolgte allerdings zufällig zu einer von vier Varianten³³.

Das Antwortverhalten der einzelnen Joker und deren objektive Validität entsprachen sich in der frequentistischen Ausprägung. Der Joker mit der objektiven Validität $v_1 = 0.92$ beantwortete die 50 Quizfragen in der Trainingsphase also insgesamt 46 mal richtig und 4 mal falsch. Die falschen Antworten verteilten sich zufällig über jeweils 10 aufeinander folgende Aufgaben in der Trainingsphase. Somit wurde gewährleistet, dass die Schwankungen der Falschantworten bei jeweils zehn aufeinander folgenden Aufgaben pro Versuchsperson minimal waren.

Nach jeweils 10 Quizfragen (Lerneinheiten) erschien ein neuer Bildschirm mit der Auflistung der vier Telefonjoker, deren Zuverlässigkeit mit einem Schieberegler auf einer Skala von 0-100% eingeschätzt werden musste (vgl. Abbildung 6.2).

³³ Um Positions- und Verwechslungseffekte zu minimieren, wurde der jeweils valideste Cue nur an 2. oder 3. Stelle und der höchst-valideste nicht in unmittelbarer Nähe zum zweitvalidesten präsentiert. Dies ergab die 4 möglichen Cue-Abfolgen: $v_3/v_1/v_4/v_2$ oder $v_4/v_1/v_3/v_2$ oder $v_2/v_3/v_1/v_4$ oder $v_2/v_4/v_1/v_3$.

Präferenzrangreihenbildung

Bitte schätze die Zuverlässigkeit von Telefonjoker A, B, C und D ein! Die vier schwarzen Dreiecke lassen sich mit der Maus wie Regler nach links und rechts verschieben. Die Einstellungen gehen von 0% Zuverlässigkeit (= der Telefonjoker hat immer falsch geantwortet) bis 100 % Zuverlässigkeit (= der Telefonjoker hat immer korrekt geantwortet). Die Einschätzungen müssen zuerst bestätigt werden, bevor es weiter gehen kann.

Telefonjoker	Geschätzte Zuverlässigkeit
A	80 % 
B	64 % 
C	92 % 
D	36 % 

Einschätzungen bestätigen

Weiter

Abbildung 6.2: Einschätzung der Zuverlässigkeiten (Validitäten) der 4 Joker in der Trainingsphase (nach jeweils 10 Lerndurchgängen) mittels Schieberegler auf einer Skala von 0-100%

Mit dieser Einschätzung wurde also sowohl die referentielle Rangordnung zwischen den einzelnen Cues als auch die subjektiv geschätzten absoluten Werte der Cue-Validitäten im Laufe der Lernphase insgesamt fünf mal gemessen. Die meisten der Versuchspersonen bewerteten allerdings jeweils die Cue-Validitäten der letzten 10 Durchgänge separat und somit unabhängig von der Vorgeschichte der vorangegangenen Durchgänge.

Die 20 Aufgaben der Anwendungsphase unterschieden sich von der Trainingsphase dadurch, dass jetzt die vier Joker vor dem definitiven Urteil aktiv um ihren Tipp angefragt werden konnten. Nach der Wahl einer der Antwortalternativen und der ersten Konfidenzangabe, konnte sich die Versuchsperson entscheiden, sich wie bisher allein auf ihr Urteil bzw. Wissen zu verlassen (Antwort ohne Zuhilfenahme eines Jokers) oder einen oder mehrere Joker um Hilfe zu bitten (Informationssuche). Die Versuchsperson war frei, welchen und wie viele Joker sie in welcher Reihenfolge befragen wollte. Beim Aufdecken eines Jokers wurde der Hinweis sichtbar, auf welche Alternative dieser tippen würde. Auch dieser Tipp konnte richtig oder falsch sein, entsprechend dem „objektiven“ Expertisegrad des Jokers.

Für das Befragen eines einzelnen Jokers wurden allerdings 400 Punkte vom Spielkonto abgezogen. Die Suchkosten wurden eingeführt, um das Aufdecken sämtlicher vier Hinweise zu verhindern und mussten auf jeden Fall bezahlt werden. Die Suchkosten waren mit $\frac{2}{5}$ des möglichen Gewinns so hoch, dass ein One-Reason decision making gefördert werden sollte. Nach Beenden der Informationssuche musste sich die Versuchsperson definitiv für eine der vier Alternativen

entscheiden und nochmals eine Konfidenzangabe vornehmen. Am Schluss der Anwendungsphase erfolgte die Erhebung einiger Personalien und ein kurzes Debriefing.

Um die Versuchspersonen zusätzlich zu motivieren, möglichst viele korrekte Entscheidungen zu treffen, wurde unter den 10 Teilnehmern mit der höchsten Gesamtpunktezahl aus Lern- und Anwendungsphase drei mal CHF 50.- verlost. Der Punktestand des Spielkontos war während des gesamten Spiels sichtbar. Vor der Trainingsphase wurde eine kurze Testphase mit drei Aufgaben und vier unterschiedlichen Jokern (als im eigentlichen Experiment) eingeführt, um das Verständnis der Versuchsperson zu testen und das Handling zu üben.

6.3 Resultate

6.3.1 Güte der gelernten Cues und ihrer Cue-Validitäten (Fragestellung 1)

Die erste Fragestellung bezog sich darauf, wie gut der Aufbau einer Repräsentation von Cue-Validitäten gelingen kann. Einerseits interessiert die Güte der absoluten Schätzung (Differenz zwischen subjektiven und ökologischen Validitätswerten) und andererseits die korrekt repräsentierte Rangfolge der Cues im Verlauf und am Ende des Trainingsdurchganges.

Als erstes soll die Abweichung der subjektiv repräsentierten Cue-Validitäten von den ökologischen Validitäten betrachtet werden. Die gemittelten Differenzbeträge zwischen den subjektiven und ökologischen Validitätswerten beträgt über sämtliche Cues und Durchgänge hinweg .20 ($SD_{Diff} = .11$). Dabei wird der objektiv 2. valideste Cue mit der geringsten Differenz am besten eingeschätzt ($M_{Diff} = .17$, $SD_{Diff} = .10$) und der objektiv valideste Cue am schlechtesten ($M_{Diff} = .23$, $SD_{Diff} = .14$). Die Art und Ausprägung der Fehleinschätzung für sämtliche vier Cues und für beide Bedingungen getrennt, zeigt Abbildung 6.3. Die ökologischen Cue-Validitäten werden massiv unterschätzt (vgl. Tabelle 6.1). Der valideste Cue ($v_1 = .92$) wird in 93.3% aller Schätzungen unterschätzt, der 2. und 3. valideste Cue mit 75.8% bzw. 79.4% und der am geringsten valideste ($v_4 = .52$) mit 71.5%.

Haben die Versuchspersonen zumindest die Rangfolge der vier Cue-Validitäten korrekt gelernt? Hier betrachten wir die Anzahl der korrekten bzw. der vertauschten Rangplätze. Das Verhältnis korrekt repräsentierter Rangplätze beträgt über sämtliche Cues und Durchgänge hinweg .46 ($SD_{Diff} = .25$) und unterscheidet sich signifikant vom Zufallsniveau von 0.25 ($t(31) = 4.78$, $p < .000$). Bei den Rangplätzen wird der objektiv am wenigsten valide Cue am häufigsten korrekt auf dem letzten Rang eingeschätzt ($M = .57$, $SD = .34$), gefolgt vom ökologisch validesten Cue mit einer korrekten Einschätzung auf Rang 1 ($M = .49$, $SD = .33$). Bei sämtlichen Versuchspersonen wird entweder der ökologisch 1. oder der 2. valideste Cue als subjektiv validester Cue, d.h. auf Rang 1, eingeschätzt (vgl. Tabelle 6.1). Die gemittelten Rangkorrelationen über die gesamte Stichprobe zeigen auf, dass die Rangfolge der Cue-Validitäten mit einigen Ausnahmen hinreichend bis sehr gut gelernt werden ($M = .72$, $SD = .34$). Bei 10 Versuchspersonen (30%) besteht eine komplette Übereinstimmung zwischen ökologischer und subjektiver Cue-Hierarchie und bei

70% der Versuchspersonen beträgt die Rangkorrelation .80 oder höher. Lediglich drei Versuchspersonen (9%) haben mit Rangkorrelationen von .00 und -.40 sehr schlecht gelernt.

Das Schätzverhalten der Versuchspersonen legt den Schluss nahe, dass diese ihre Cue-Einschätzungen lediglich auf die jeweils 10 zurückliegenden Lerndurchgänge bezogen haben. Deshalb ist der Mittelwert über alle fünf Einzelschätzungen der einzig valide Prädiktor für die Lerngüte.

6.3.2 Kein Unterschied in den Repräsentationen zwischen den beiden Bedingungen

Sowohl bei den absoluten Schätzwerten (Differenzbeträge) wie auch bei den Rangeinschätzungen sind anhand der grossen Standardabweichungen grosse interindividuelle Unterschiede in der Lerngüte zu erkennen. Ein möglicher Erklärungsfaktor ist der Einfluss der Ähnlichkeit von zwei benachbarten ökologischen Cue-Validitäten. Objektiv nahe beieinanderliegende Cue-Validitäten sollten wegen einer möglichen Interferenz schlechter gelernt werden. Um diese Hypothese zu testen, wurden die Versuchspersonen vorgängig zufällig einer von zwei Bedingungen zugeteilt: 16 Personen (Bedingung A) lernten anhand der ökologischen Validitäten $v_1 = 0.92$, $v_2 = 0.76$, $v_3 = 0.60$ und $v_4 = 0.52$, d. h. mit einer Differenz von $v_{diffA1-2} = 0.16$ zwischen dem 1. und 2. validesten Cue. 16 Personen (Bedingung B) lernten mit einem geringeren Unterschied von $v_{diffB1-2} = 0.08$ zwischen dem 1. und 2. validesten Cue ($v_1 = 0.92$, $v_2 = 0.84$, $v_3 = 0.68$ und $v_4 = 0.52$).

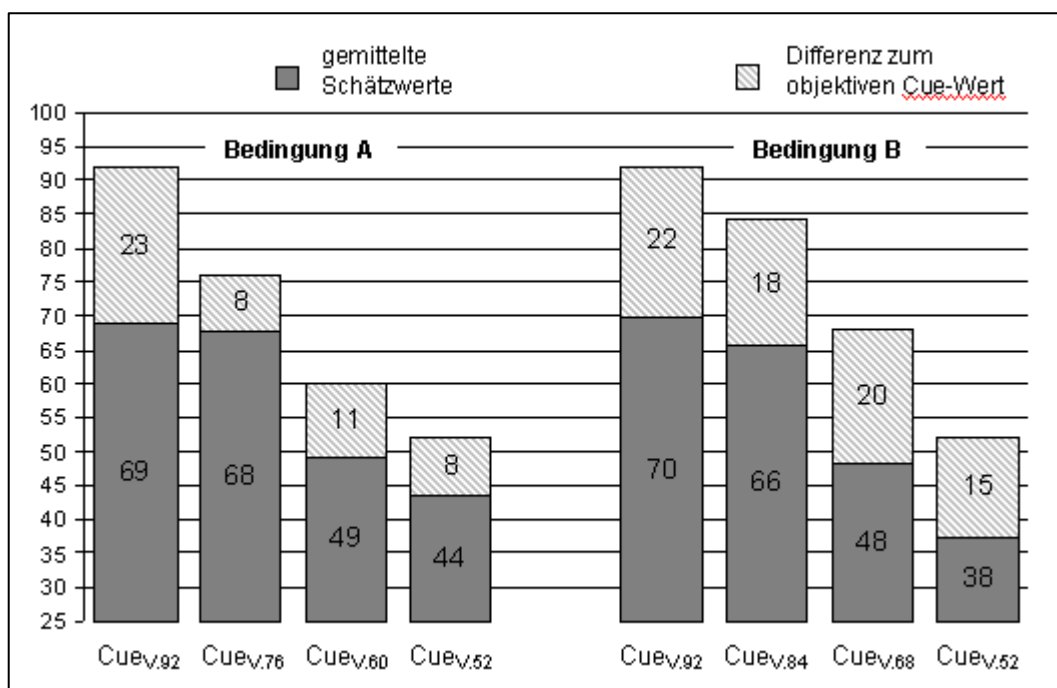


Abbildung 6.3: Mittlere Einschätzungen der subjektiv repräsentierten Cue-Validitäten in % (dunkelgrau) getrennt nach den beiden Bedingungen A und B und ihre Abweichungen (Unterschätzungen) zu den wahren Validitäts-Werten (Differenz in hellgrau).

Bezogen auf die Lerngüte unterscheiden sich die Personen in beiden Bedingungen statistisch nicht überzufällig. Personen in Bedingung B weisen zwar im Schnitt beim objektiv 2. und 3.

validesten Cue etwas höhere aber keine signifikant unterschiedlichen Differenzbeträge auf (vgl. Abbildung 6.3). Dies gilt ebenso für das Verhältnis der vertauschten Rangplätze. Wird in Bedingung B der ökologisch valideste Cue etwas weniger und der am wenigsten valide Cue etwas mehr vertauscht, so ist die Vertauschungsrate der beiden mittleren Cues in beiden Bedingung praktisch identisch. Die Rangkorrelationen zwischen der Bedingung A ($M = .66$, $SD = .42$) und B ($M = .77$, $SD = .23$) unterscheiden sich nicht überzufällig ($t(30) = 0.91$, $p = .19$, einseitig). Da kein Unterschied zwischen den Bedingungen besteht, werden die beiden Versuchspersonengruppen für die weiteren präsentierten Resultate zusammengelegt und als Einheit betrachtet.

Tabelle 6.1: Subjektiv (S) repräsentierte Cue-Validitäten als gemittelte (M) Einschätzungen pro Cue (MSC_{V1-4}) über fünf Messzeitpunkte während der Lerndurchphase. Die Validität von Cue_{V1} entspricht .92, Cue_{V2} .76 (für Bedingung A) oder .84 (B), etc. $MSCV_{max}$ ist die subjektiv valideste Cue-Einschätzung pro Versuchsperson, $MSCV_{Diffmax-2}$ die Differenz zum subjektiv zweitvalidesten. Die Cue-Hierarchie bildet die objektiven Cue-Validitäten in der Reihenfolge der subjektiv repräsentierten Mittelwerte ab, wobei $1 > 2 > 3 > 4$ die 100% korrekt repräsentierte Abfolge darstellt.

Vp-No.	Bed	MSC_{V1}	MSC_{V2}	MSC_{V3}	MSC_{V4}	$MSCV_{max}$	$MSCV_{Diffmax-2}$	Cue-Hierarchie	Typ
29	B	.90	.64	.55	.13	.90	.27	$1 > 2 > 3 > 4$	ORDM
32	A	.87	.71	.58	.52	.87	.16	$1 > 2 > 3 > 4$	ORDM
6	B	.87	.63	.49	.39	.87	.23	$1 > 2 > 3 > 4$	ORDM
18	A	.83	.66	.39	.30	.83	.16	$1 > 2 > 3 > 4$	ORDM
9	A	.82	.58	.46	.28	.82	.24	$1 > 2 > 3 > 4$	ORDM
26	A	.76	.66	.54	.43	.76	.10	$1 > 2 > 3 > 4$	OR&MRDM
20	A	.80	.52	.47	.34	.80	.29	$1 > 2 \sim 3 > 4$	ORDM
27	B	.70	.63	.62	.58	.70	.07	$1 > 2 \sim 3 \sim 4$	ORDM
22	A	.68	.64	.60	.04	.68	.04	$1 \sim 2 \sim 3 > 4$	OR&MRDM
2	A	.65	.63	.49	.22	.65	.02	$1 \sim 2 > 3 > 4$	ORDM
23	B	.71	.71	.50	.26	.71	.00	$1 = 2 > 3 > 4$	ORDM
31	B	.86	.58	.43	.45	.86	.28	$1 > 2 > 4 \sim 3$	ORDM
1	B	.78	.66	.34	.40	.78	.12	$1 > 2 > 4 > 3$	ORDM
10	B	.87	.37	.76	.30	.87	.12	$1 > 3 > 2 > 4$	ORDM
5	A	.86	.80	.75	.80	.86	.06	$1 > 4 \sim 2 \sim 3$	ORDM
11	A	.54	.47	.35	.51	.54	.04	$1 \sim 4 \sim 2 > 3$	OR&MRDM
19	B	.80	.88	.72	.64	.88	.08	$2 > 1 > 3 > 4$	ORDM
21	B	.63	.82	.48	.27	.82	.18	$2 > 1 > 3 > 4$	OR&MRDM
7	B	.75	.81	.62	.59	.81	.06	$2 > 1 > 3 \sim 4$	OR&MRDM
8	A	.77	.78	.62	.36	.78	.00	$2 \sim 1 > 3 > 4$	OR&MRDM
15	B	.64	.69	.58	.45	.69	.05	$2 \sim 1 > 3 > 4$	OR&MRDM
13	B	.65	.66	.51	.19	.66	.01	$2 \sim 1 > 3 > 4$	OR&MRDM
16	A	.61	.62	.41	.35	.62	.01	$2 \sim 1 > 3 > 4$	OR&MRDM
4	B	.43	.48	.37	.15	.48	.05	$2 \sim 1 > 3 > 4$	OR&MRDM
28	A	.71	.85	.60	.65	.85	.14	$2 > 1 > 4 \sim 3$	ORDM
12	A	.73	.81	.27	.65	.81	.08	$2 > 1 > 4 > 3$	ORDM
17	B	.62	.69	.22	.30	.69	.08	$2 > 1 > 4 > 3$	ORDM
3	B	.41	.52	.23	.36	.52	.11	$2 > 1 > 4 > 3$	ORDM
30	A	.54	.71	.58	.35	.71	.13	$2 > 3 \sim 1 > 4$	OR&MRDM
25	B	.54	.73	.28	.54	.73	.19	$2 > 4 \sim 1 > 3$	OR&MRDM
24	A	.39	.64	.26	.44	.64	.20	$2 > 4 \sim 1 > 3$	OR&MRDM
14	A	.47	.79	.50	.73	.79	.06	$2 > 4 > 3 \sim 1$	OR&MRDM

6.3.3 Bewertung der subjektiv repräsentierten Cue-Validitäten

Die absoluten Cue-Validitäten werden schlecht repräsentiert und liegen im Mittel um .20 neben dem wahren Validitäts-Wert und zumeist zu tief. Auf der anderen Seite scheint die Mehrheit der Versuchspersonen aber in der Lage zu sein, auch innert kurzer Zeit, d.h. innerhalb von 50 Lern-durchgängen, die ökologische Cue-Hierarchie annähernd gut bis perfekt zu erfassen. Dabei bestehen grosse interindividuelle Unterschiede. Immerhin ist der subjektiv als am validesten eingeschätzte Cue bei sämtlichen Versuchspersonen entweder der ökologisch valideste oder zweitvalideste Cue (vgl. Tabelle 6.1). Entscheidend für die Umsetzung der subjektiv repräsentierten Cues ist allerdings nicht die Adäquatheit zu den realen ökologischen Validitäten, sondern dass überhaupt eine interne Cue-Rangordnung bezüglich der Validitäten aufgebaut wird. Auch mit einer suboptimalen Cue-Repräsentation lässt sich die Take The Best-Heuristik (TTB) anwenden. Für die Anwendung von TTB wird lediglich eine stabile relationale Hierarchie der subjektiv repräsentierten Cue-Validitäten vorausgesetzt, unabhängig von der wahren Ausprägung der Cue-Werte. Die Mehrzahl der Versuchspersonen hat über alle fünf Einzelschätzungen hinweg eine eindeutige und stabile subjektive Rangfolge der vier Cues bilden können. Die Auswirkungen des individuellen Lernens können nun auf die Anwendung von TTB hin untersucht werden. Die Art der beobachteten Cue-Repräsentation reicht allerdings nicht für die Anwendung eines kompensatorisch gewichteten Entscheidungsmodells wie etwa Franklin's Rule oder einer multiplen Regression (vgl. Hausmann & Läge, 2005a), denn die Gewichtung benötigt adäquate absolute Cue-Repräsentationen. Neben TTB sind die Bedingungen hingegen auch für ein kompensatorisch ungewichtetes Entscheidungsmodell wie z.B. Dawes' Rule erfüllt.

6.3.4 Einfluss der wahrgenommenen Cue-Güte auf die Anwendung von TTB (Fragestellung 2)

In den 20 Aufgaben der Anwendungsphase haben sämtliche Personen die Verfügbarkeit der vier Joker benutzt. In durchschnittlich 21% der Aufgaben ($M = 4.16$, $SD = 3.55$) haben die Versuchspersonen ohne Zuhilfenahme eines Jokers geantwortet, z.B. weil sie die korrekte Antwort der Quizfrage schon gewusst haben ($n = 133$). Im folgenden werden nur noch diejenigen Aufgaben für die Auswertung berücksichtigt, in denen mindestens ein Joker befragt wurde ($n = 507$). Diese Aufgaben mit Informationssuche lassen sich darin unterscheiden, ob genau eine oder aber mehrere Informationen gesucht wurde. Im ersten Fall sprechen wir von einer non-kompensatorischen Informationsverarbeitung bzw. von One-Reason Decision Making (ORDM), im zweiten Fall von einer kompensatorischen bzw. von More-Reason Decision Making (MRDM). Da sämtliche Cues zu 100% diskriminieren, d.h. immer auf eine Antwortalternative hinweisen, ist das Anwenden eines ORDM identisch mit der korrekten Anwendung der Stopppregel der Take The Best-Heuristik (TTB).

In 77% der Aufgaben mit Informationssuche verfolgten die Versuchspersonen ein ORDM ($n = 392$). In 4% der Aufgaben mit ORDM wird die Entscheidungsregel von TTB verletzt, d.h. die Person entscheidet sich gegen die Alternative, für die der Cue sprechen würde. Dieser Anteil ist vergleichbar mit früheren Experimenten (vgl. Läge, Daub, Christen & Hausmann, 2005) und kann als Fehlerwert vernachlässigt werden. Die Suchregel gemäss TTB wird in 78% der Aufgaben mit ORDM korrekt angewendet. Für die restlichen Fälle bedeutet dies, dass nicht der Cue

aufgedeckt wurde, den die Person im Mittel über die fünf Cue-Einschätzungen nach je 10 Lerndurchgängen als subjektiv validesten repräsentiert hat. Hier gibt es viele Personen, die systematisch entweder keine oder aber viele Verstösse gegen die Suchregel aufweisen.

Die 23% der Aufgaben, die gemäss einem MRDM gelöst wurden ($n = 115$), verteilen sich auf 85% mit zwei, 8% mit drei und 7 % mit vier gesuchten Informationen.

6.3.5 Anwendung der Cue-Repräsentationen im Sinne von TTB

Nutzen die Versuchspersonen die wahrgenommene Cue-Güte (subjektive Validitäten) in der Lernphase im Sinne der Take The Best-Heuristik (TTB)? Für die Cue-Güte sprechen zwei unabhängige Variablen, welche das Ausmass an korrektem TTB-Verhalten (als abhängige Variable) beeinflussen könnten. Einerseits bestimmt die wahrgenommene Differenz zwischen 1. und 2. validesten Cue die korrekte Anwendung der Suchregel von TTB (vgl. Luce, 1956) und andererseits bestimmt die Höhe der Cue-Validität des subjektiv validesten Cues die subjektive Urteilssicherheit beim Beenden der Informationssuche (vgl. Chaiken, Liberman & Eagly (1989); Eagly & Chaiken (1993); Hausmann & Läge (2005c)). Die Produkt-Moment-Korrelationen zwischen den beiden Variablen und dem Anteil korrekt angewendeter TTB-Regeln zeigen auch tatsächlich einen starken positiven Zusammenhang. Für die Suchregel wurde für jede Versuchsperson ($n = 32$) die Anzahl der Aufgaben, in denen der subjektiv valideste Cue zuerst gesucht wurde ins Verhältnis gesetzt zur Anzahl der Aufgaben, in denen genau ein Cue gesucht wurde (ORDM). Dieses Verhältnis korreliert signifikant mit der Höhe der Differenz zwischen dem 1. und 2. validesten Cue ($r = .45$, $p = .01$). Für die Stopppregel wurde für jede Versuchsperson das Verhältnis der Anzahl Aufgaben mit ORDM zur Gesamtzahl der Aufgaben mit Informationssuche gebildet (ORDM und MRDM). Dieses Verhältnis weist ebenfalls eine hohe, aber keine signifikante Korrelation auf mit der Höhe der Cue-Validität des subjektiv validesten Cues ($r = .31$, $p = .08$). Beide Variablen haben einen Einfluss auf die Anwendung der TTB-Strategie.

6.3.6 Individuelles Entscheidungsverhalten

Wenn wir nun die Frage des Einflusses der wahrgenommenen Cue-Güte auf die korrekte Anwendung von TTB beantworten wollen, können wir dies auch auf dem individuellen Versuchspersonenniveau tun. Dazu müssen wir die Personen in TTB-User bzw. Non-TTB-User einteilen.

Die nachfolgende Klassifikation bezieht sich primär auf die korrekte Anwendung der Stopppregel, also auf die Frage, ob die Versuchsperson ein ORDM anwendet. Die Auswertung zeigt, dass sich die Versuchspersonen tatsächlich im Umfang der gesuchten Information unterscheiden. 18 Personen (56%) lassen sich als konsequente One-Reason decision maker (ORDM) bezeichnen. Als Klassifikationskriterium diente ein einfacher Binomialtest der Anzahl Aufgaben mit ORDM im Verhältnis zu Aufgaben mit MRDM. Der Anteil an ORDM liegt bei dieser Gruppe deshalb entsprechend hoch bei $M = .98$ ($SD = .04$). Die Gruppe ORDM zeichnet sich zudem dadurch aus, dass sie (neben dem gelegentlichen Raten) zu 89.5% exakt immer dieselbe Such- und Entscheidungsstrategie verfolgen. 9 Personen wenden sogar zu 100% immer dieselbe Such- und Stopppregel an.

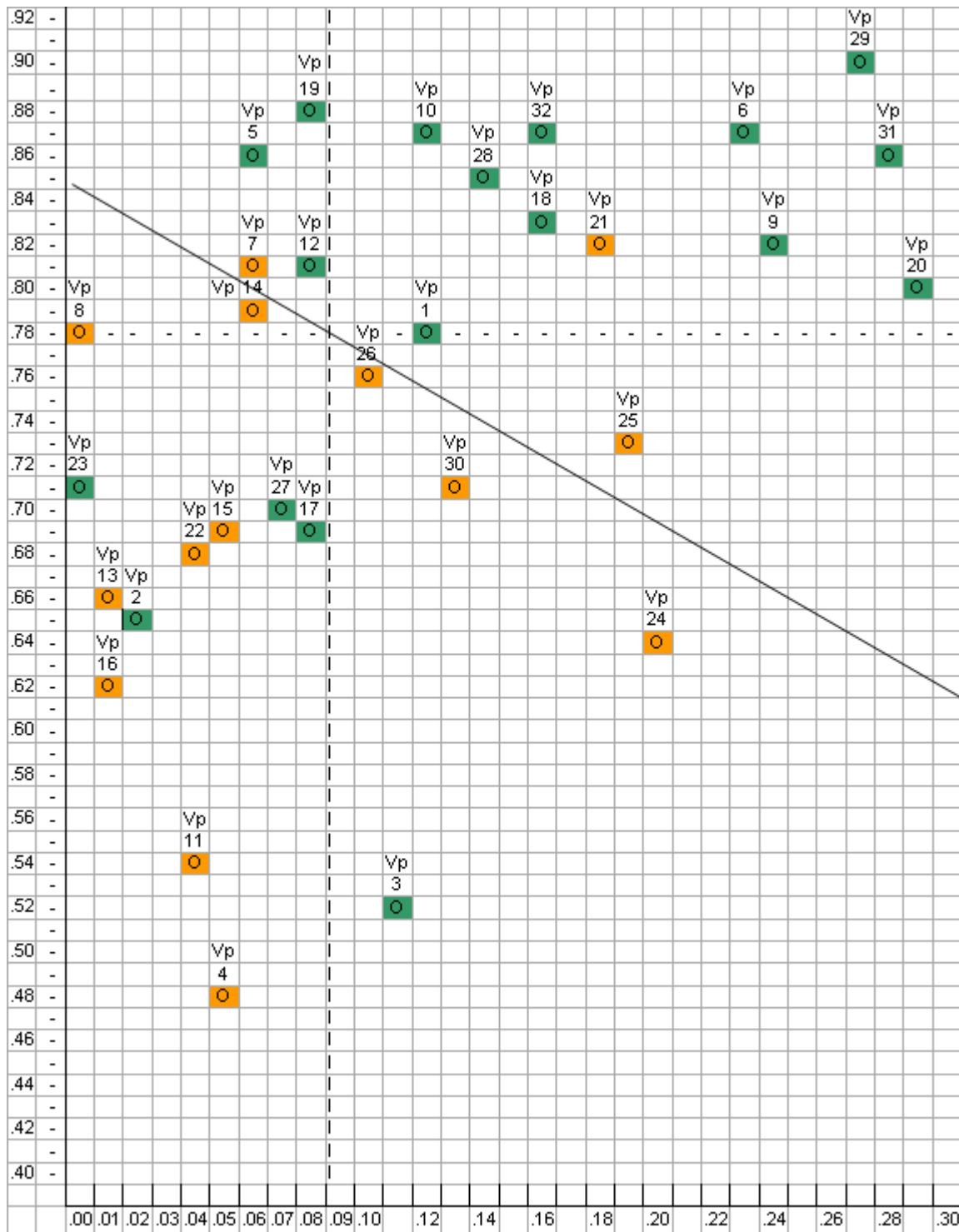


Abbildung 6.4: Die 32 Versuchspersonen abgetragen nach der gemittelten Validität des subjektiv als am validesten repräsentierten Cues (vertikale Achse von .40 bis .92) und die Differenz der Validitäten zwischen dem subjektiv validesten und dem zweitvalidesten Cue (horizontale Achse von .00 bis .30). Man erkennt, dass Personen mit konsequentem ORDM (grün) durch höhere Validitäten und grössere Differenzen auffallen. Personen mit vermehrtem MRDM (OR&MRDM, orange) hingegen weisen tendenziell weniger valide Cues auf und eine geringe Differenz der beiden validesten Cues.

Die restlichen 14 Personen zeichnen sich dadurch aus, dass sie sowohl ein One-Reason decision making als auch ein More-Reason decision making zeigen. Wir nennen sie die Gruppe der

OR&MRDM. Das Verhältnis an reinem ORDM im Vergleich zu sämtlichen Aufgaben, in denen Informationen gesucht wurden, beträgt innerhalb dieser Gruppe im Schnitt 53% ($SD = 14\%$).

Der Zusammenhang zwischen Entscheidungstyp und den beiden unabhängigen Variablen der Lerngüte ist in Abbildung 6.4 abgetragen. Die Produkt-Moment-Korrelation zwischen den beiden Variablen beträgt $r = .48$ ($p < 0.01$). Sämtliche Versuchspersonen sind mit ihren Werten auf den beiden Achsen „Höhe des subjektiv validesten Cues“ und „Differenz zwischen dem subjektiv validesten und 2. validesten Cue“ eingetragen. Die Gruppe mit reinem ORDM hat den subjektiv validesten Cue signifikant höher repräsentiert als die Gruppe OR&MRDM ($t(30) = 2.77$, $p = .004$, einseitig). Zudem ist die Differenz zwischen dem subjektiv 1. und 2. validesten Cue bei der Gruppe mit reinem ORDM signifikant höher ($t(30) = 2.06$, $p = .025$, einseitig).

6.4 Diskussion

Mit der Take The Best-Heuristik (TTB) haben Gigerenzer, Hoffrage und Kleinbölting (1991) im Rahmen ihrer Theorie Probabilistischer Mentaler Modelle (PMM) eine attraktive Urteilsheuristik postuliert. In Entscheidungssituationen unter Unsicherheit trifft sie auf eine einfache und schnelle Art gute Entscheidungen (simple, fast and frugal) (vgl. Hausmann & Läge, 2005a). Dass TTB in Computersimulationen eine ebenso gute Performance erzielt wie aufwändigere kompensatorische Entscheidungsmodelle schenkte ihr in der Entscheidungsforschung viel Sympathie und Beachtung. Im empirischen Test lässt TTB allerdings viele Fragen offen (vgl. Kritik von Bröder, 2000a; 2004; Hausmann, 2004). Obwohl TTB-Entscheidungsverhalten bei Versuchspersonen in Laborexperimenten nachgewiesen werden konnte, tun dies nicht alle, wenige machen dies konsequent und generell nur unter bestimmten einschränkenden Bedingungen (vgl. Hausmann & Läge, 2005a). Zudem ist eine wichtige Frage bislang unbeantwortet geblieben: Setzen Personen subjektiv repräsentiertes Wissen über Cue-Validitäten erfolgreich, d.h. im Sinne von „take the best Cue“ überhaupt ein? Dies ist abhängig davon, wie gut Personen die ökologischen Cue-Validitäten intern repräsentiert haben. Nur mit einem realistisch aufgebauten Bild der Cue-Validitäten lassen sich über eine längere Zeit hinweg Entscheidungsaufgaben derselben Art erfolgreich lösen. Um allerdings die relevanten ökologischen Cues effektiv lernen zu können, müssen folgende drei Bedingungen erfüllt sein: Cues müssen beobachtbar sein, eine hohe Validität besitzen und zwischen den Alternativen möglichst oft diskriminieren. Sämtliche der Bedingungen wurden im vorliegenden Experiment berücksichtigt. Die Diskriminationsrate wurde unter Berücksichtigung von personifizierten Cues sogar auf 100% gesetzt.

Die subjektiven Einschätzungen der Cue-Validitäten in der Lernphase haben gezeigt, dass die meisten der Personen grundsätzlich bereits schon nach wenigen Beobachtungen in der Lage sind, eine stabile Rangfolge der Cues aufzubauen, was als Anforderung für die Anwendung von TTB hinreichend ist. Die Voraussetzungen für die Anwendung einer Heuristik im Sinne von TTB war für praktisch sämtliche der Personen nach spätestens 50 Lerndurchgängen gegeben. Weshalb die einen Personen besser lernen als die anderen muss offen gelassen werden.

Für die Anwendung von TTB ist es nicht wesentlich, ob die „objektiven“ Cue-Validitäten der ökologischen Umwelt zu 100% korrekt repräsentiert werden. Viel wichtiger ist es, dass intern eine stabile Cue-Hierarchie aufgebaut werden kann. Das konkrete Entscheidungsverhalten wird von mindestens zwei Faktoren gesteuert, die die Anwendung gemäss TTB einschränken können: die wahrgenommene Validitäts-Differenz zwischen den Cues und die absolute Höhe der repräsentierten Cue-Validitäten in Bezug auf das individuelle Anspruchsniveau an Urteilssicherheit an die konkrete Entscheidungssituation.

Bei Personen, die die beiden subjektiv validesten Cues als ähnlich valide repräsentiert haben, führt dies zu einer erhöhten Vertauschbarkeit in der Suchfolge und somit zu einer Verletzung der TTB-Suchregel. Die Cues werden nicht mehr in rein lexikographischer Abfolge abgefragt, sondern semilexikographisch, d.h. in einer abwechselnden Reihenfolge (vgl. Luce, 1956). Wie gross diese repräsentierte Differenz sein muss, um als eigenständige Cues wahrgenommen zu werden, kann an dieser Stelle nicht beantwortet werden und bleibt zurzeit eine offene Forschungsfrage.

Ein substanziellerer Einwand gegen TTB betrifft die Verletzung der Stopppregel. Ein frühzeitiger Abbruch der Informationssuche könnte noch relativ bescheiden mit den Spezifikationen der Entscheidungssituation erklärt werden, z.B. mit hohen Suchkosten (vgl. Hausmann, Christen & Läge, 2005). Eine Missachtung der Stopppregel in die andere Richtung ist problematischer, da eine zusätzliche Informationssuche auch dadurch erklärt werden könnte, dass ein anderes Entscheidungsmodell bevorzugt wird und die Person z. B. ein kompensatorisches More-Reason Decision Making (vgl. Hausmann, 2004) dem non-kompensatorischen One-Reason Decision Making vorzieht.

Die Resultate haben gezeigt, dass die Höhe des subjektiv validesten Cues die konsequente Verwendung von TTB bestimmt. Cues, welche subjektiv als hoch valide eingeschätzt werden, haben ein konsequentes ORDM zur Folge. Bei Personen, welche den subjektiv validesten Cue geringer einschätzen, zeigen häufiger eine Verletzungen der Stopppregel. Chaiken, Liberman & Eagly (1989) haben mit dem Konzept der gewünschten Urteilssicherheit einen individuellen Einflussfaktor vorgeschlagen, der das Verfolgen einer spezifischen Entscheidungsstrategie (Entscheidungsregel) gänzlich überflüssig machen könnte. Grundsätzlich werden Personen nur solange nach Informationen suchen, bis sie ein hinreichend sicheres Urteil fällen können. Eine Entscheidung wird dann möglich, wenn die aktuelle Urteilssicherheit der Informationslage dem gewünschten Niveau entspricht oder dieses übersteigt. Auf das Konzept der Cue-Validitäten übertragen würde dies bedeuten, dass Personen sich anhand derjenigen Cues entscheiden, deren Validität dieses individuelle Anspruchsniveau übersteigt. Hausmann & Läge (2005c) konnten zeigen, dass ein solches individuelles Niveau existiert und bei vier Alternativen um die .75 liegt. Es kann also sehr gut sein, dass die Gruppe ORDM den subjektiv validesten Cue als so hoch valide repräsentiert hat, dass er ihr Anspruchsniveau an gewünschter Urteilssicherheit (DLC) erfüllt. Die Validität des subjektiv validesten Cues liegt bei dieser Gruppe im Schnitt auch tatsächlich über .75 ($M = .79$, $SD = .10$). Dies ist signifikant höher als eine Gleichverteilung von .75 ($t(17) = 1.80$, $p = .043$ (one-tailed)).

Die Gruppe OR&MRDM zeichnet sich hingegen dadurch aus, dass ihre subjektiv valideste Cue-repräsentation unterhalb von .75 liegt ($M = .69$, $SD = .10$). Dies ist signifikant tiefer als eine Gleichverteilung von .75 ($t(13) = 2.10$, $p = .025$ (one-tailed)). Somit genügt dieser eine Cue nicht, um eine befriedigende Antwort mit entsprechend hoher Urteilssicherheit zu erhalten. Weshalb diese Gruppe allerdings trotzdem ab und zu ORDM zeigt, kann durch das Versuchssetting bzw. das Vorwissen erklärt werden. Da es sich bei den Aufgaben um reale Quizfragen handelte, kann vermutet werden, dass mindestens eine der Antwortalternativen von Beginn weg präferiert wurde. Diese Präferenz kann als eigenständiger interner Hinweis-Cue (Knowledge-based Preference Cue) betrachtet werden.

Liegt eine klare Präferenz für eine der Alternativen vor und ist die subjektive Urteilssicherheit des Preference-Cues genügend gross, kann die entsprechende Alternative ohne weitere (interne oder externe) Informationssuche gewählt werden (die Antwort kann als „gewusst“ oder „ziemlich sicher“ abgebucht werden).

Liegt die Ausprägung der Präferenz unterhalb der individuellen gewünschten Sicherheitsschwelle, muss zusätzliche Probability-Information beschafft werden. Zeigt die Ausprägung eines Probability-Cues in die selbe Richtung wie die des Präferenz-Cues, kann die Alternative ebenfalls problemlos gewählt werden (konfirmatorische Information). Liegt die subjektive Validität des Probability-Cues zudem oberhalb der gewünschten Sicherheitsschwelle, kann sie auch eine vorgängige Präferenz umpolen. Von aussen betrachtet zeigt die Versuchsperson in beiden Fällen ein ORDM.

Etwas komplizierter wird es, wenn sowohl der Preference-Cue als auch der erstgesuchte Probability-Cue unterschiedliche (non-konfirmatorische) Hinweise geben und beide unterhalb der Sicherheitsschwelle liegen. Hier muss ein zweiter externer Probability-Cue gesucht werden, der einen der beiden Hinweise bestätigen könnte. In diesem Fall erfüllt keiner der Cues das Anspruchsniveau an Urteilssicherheit, aber die Gesamtwahrscheinlichkeit wird dadurch erhöht, dass nun zwei Hinweise für dieselbe Alternative sprechen. Hier könnte man von einem ungewichteten Tallying sprechen (vgl. Hausmann & Läge, 2005a), wobei mindestens zwei Hinweise in die gleiche Richtung einen anderen Hinweis in eine andere übertrumpfen müssen.

Gibt es dazu Hinweise in den Daten? Die Höhe dieser Präferenz haben wir jeweils nach der Präsentation der Quizfrage gemessen, indem wir ein Konfidenzurteil abgefragt haben. Wenn die Konfidenz 100% beträgt (absolut sicher), müssten sich die Personen, wenn sie sich rational verhalten würden, allein auf diese Urteilssicherheit stützen und würden keine externen Informationen benötigen. Dies trifft auch in allen entsprechenden Aufgaben ($n = 31$) mit der Konfidenzangabe 100% zu. Dasselbe gilt auch für 94% der Aufgaben mit der Konfidenzangabe 91%-99% (ziemlich sicher) ($n = 17$). Liegt die Konfidenz unterhalb von 90%, dann werden überwiegend externe Informationen gesucht. Für die Konfidenzangabe 61%-90% (eher sicher) wird in 36% auf eine externe Informationssuche verzichtet ($n = 53$), für 26%-60% (ziemlich unsicher) sind es noch 18% ($n = 192$) und für die Konfidenzangabe 25% (völlig geraten) nur noch 9% ($n = 347$).

6.4.1 Implikationen für die Plausibilität von TTB

Die Take The Best-Heuristik (TTB) muss als Spezialfall bezeichnet werden. Eine optimale Anwendung von TTB ist dann zu erwarten, wenn Personen über keine Präferenz in eine der Alternativen verfügen, intern eine klare Rangfolge der Cues repräsentiert haben (genügende Cue-Differenz) und sämtliche Cue-Validitäten dem individuellen Anspruchsniveau an gewünschter Urteilssicherheit (DLC) entsprechen. Zudem muss eine Beschränkung bei der Informationssuche vorliegen, wobei z.B. die externe Informationssuche mit Suchkosten verknüpft sein kann, die aber nicht zu hoch oder zu tief liegen dürfen (vgl. Hausmann, Christen & Läge, 2005). Bei dieser Fülle von Einschränkungen muss überprüft werden, ob ein generelleres Entscheidungsmodell, bei welchem das Sicherheitsdefizit an Urteilssicherheit den Umfang an gesuchter Information steuert (vgl. Hausmann & Läge, 2005c), das Entscheidungsverhalten der Personen bei Entscheidungen unter Unsicherheit nicht besser und sparsamer erklären kann als TTB.

Das vorliegende Experiment kann die *interindividuelle* Varianz in der Anwendung von TTB bzw. ORDM mit der unterschiedlichen Repräsentation der Cue-Validitäten erklären, während Hausmann & Läge (2005c) aufzeigen konnten, wie das Konzept der gewünschten Urteilssicherheit die *intraindividuelle* Varianz erzeugt.

Welche Schlussfolgerungen ergeben sich bezüglich des Erlernens von Probability-Cues für das experimentelle Setting beim Nachweis von TTB? Die vorliegende Untersuchung hat klar gezeigt, dass die Art und Weise der subjektiv repräsentierten Cues das Such- und Entscheidungsverhalten einer Person bestimmt. Das vorgängige Cue-Lernen ist also ein notwendiger Faktor für eine nachfolgende gedächtnisbasierte Informationssuche. Ein Cue-Learning ist aber experimentell nicht unproblematisch. Einerseits könnte man Versuchspersonen so lange lernen lassen, bis sie die gewünschten Lernkriterien (Ranghierarchie der Cues und absolute Höhe der Cue-Validität) korrekt repräsentiert haben. Andererseits könnte man sich konsequent auf die subjektiv wahrgenommenen Cue-Hierarchie und den Cue-Ausprägungen stützen. Jedes Entscheidungsexperiment müsste dann wie im vorliegenden Experiment mit den individuellen Werten (subjektive Einschätzungen und Gewichtungen) einer Versuchsperson durchgeführt und ausgewertet werden. Jede Versuchsperson müsste als eigene Versuchsbedingung klassifiziert werden und ein fairer Vergleich über das einzelne Individuum hinweg ist somit kaum mehr möglich. Für einen Nachweis von TTB scheint es daher fairer, den „Stör-“, bzw. Einflussfaktor“ Cue-Learning experimentell zu kontrollieren, indem den Probanden die ökologischen Werte von Anfang an vorgegeben werden.

6.5 Literatur

- Atneave, F. (1953). Psychological probability as a function of experienced frequency. *Journal of Experimental Psychology*, 46, 81–86.
- Brase, G. L., Cosmides, L., & Tooby, J. (1998). Individuation, Counting, and statistical inference: The role of frequency and whole-object representations in judgment under uncertainty. *Journal of Experimental Psychology: General*, 127, 3–21.

- Bröder, A. (2000a). „Take The Best – Ignore The Rest”: Wann entscheiden Menschen begrenzt rational? [“Take The Best - Ignore The Rest” When are people's decisions boundedly rational?]. Lengerich: Pabst Science Publishers.
- Bröder, A. (2000b). A methodological comment on behavioral decision research. *Psychologische Beiträge*, 42, 645–662.
- Bröder, A. (2000c). Assessing the empirical validity of the “Take The Best”-heuristic as a model of human probabilistic inference. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 26, 1332–1346.
- Bröder, A. (2002). Take The Best, Dawes' Rule, and compensatory decision strategies: A regression-based classification method. *Quality & Quantity*, 36, 219–238.
- Bröder, A. (2003). Decision making with the “adaptive toolbox”: Influence of environmental structure, intelligence, and working memory load. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 29, 611–625.
- Bröder, A. (2005). *Entscheiden mit der „Adaptiven Werkzeugkiste”: Ein empirisches Forschungsprogramm*. Lengerich: Pabst Science Publishers.
- Bröder, A., & Schiffer, S. (2003). “Take The Best” versus simultaneous feature matching: Probabilistic inferences from memory and effects of representation format. *Journal of Experimental Psychology: General*, 132, 277–293.
- Brunswik, E. (1952). The conceptual framework of psychology. In *International Encyclopedia of Unified Science*, 1, 10. Chicago, IL: The University of Chicago Press.
- Chaiken, S., Liberman, A., & Eagly, A. H. (1989). Heuristic and systematic information processing within and beyond the persuasion process. In J. S. Uleman, & J. A. Bargh (Eds.), *Unintended thought* (pp. 212–252). New York: Guilford Press.
- Chater, N., Oaksford, M., Nakisa, R., & Redington, M. (2003). Fast, frugal, and rational: How rational norms explain behavior. *Organizational Behavior and Human Decision Processes*, 90, 63–86.
- Czerlinski, J., Gigerenzer, G., & Goldstein, D. G. (1999). How good are simple heuristics? In G. Gigerenzer, P. M. Todd & the ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 97–118). New York: Oxford University Press.
- Eagly, A. H., & Chaiken, S. (1993). *The Psychology of Attitudes*. San Diego, CA: Harcourt Brace Jovanovick.
- Erdbauer, M. (Hrsg.) (2001). *Wissensportal Allgemeinbildung: Das aktuelle Wissen unserer Zeit in Frage und Antwort*. Trautwein Lexikon-Edition. München: Compact.
- Estes, W. K. (1976). The cognitive side of probability learning. *Psychological Review*, 83, 37–64.
- Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, 103, 650–669.
- Gigerenzer, G., & Goldstein, D. G. (1999). Betting on one good reason: The Take The Best heuristic. In G. Gigerenzer, P. M. Todd & the ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 75–95). New York: Oxford University Press.
- Gigerenzer, G., & Selten, R. (Eds.) (2001). *Bounded rationality: The adaptive toolbox. Dahlem Workshop report*. Cambridge, MA: The MIT Press.
- Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic Mental Models: A Brunswikan theory of confidence. *Psychological Review*, 98, 506–528.
- Gigerenzer, G., Todd, P. M., & the ABC Research Group (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.
- Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General*, 108, 356–388.

- Hasher, L., & Zacks, R. T. (1984). Automatic processing of fundamental information. *American Psychologist*, 39, 1372–1388.
- Hausmann, D. (2004). *Informationssuche im Entscheidungsprozess. Die Nützlichkeit von Hinweis-Cues und der Anspruch an Urteilssicherheit*. Dissertationsschrift. Zürich: Zentralstelle der Studentendruckerei.
- Hausmann, D., & Läge, D. (2005a). *Take The Best im Test: Empirischer Stellenwert einer einfachen Urteilsheuristik und ein alternatives Kriteriumsmodell*. AKZ-Forschungsbericht 03. Zürich: Angewandte Kognitionspsychologie.
- Hausmann, D., & Läge, D. (2005b). *Wenn Limiten zum Strategiewechsel führen: Wechsel der Suchstrategie unter situativen Einschränkungen*. AKZ-Forschungsbericht 06. Zürich: Angewandte Kognitionspsychologie.
- Hausmann, D., & Läge, D. (2005c). *Der Wunsch nach Sicherheit: Wie das Anspruchsniveau an Urteilssicherheit in Entscheidungssituationen unter Unsicherheit das Informationssuchverhalten steuert*. AKZ-Forschungsbericht 09. Zürich: Angewandte Kognitionspsychologie.
- Hausmann, D., Christen, S., & Läge, D. (2005). *Wie viel bezahlen für eine valide Information? Suchkosten als limitierender Faktor der Informationssuche*. AKZ-Forschungsbericht 07. Zürich: Angewandte Kognitionspsychologie.
- Läge, D., Daub, S., Christen, S., & Hausmann, D. (2005). *Was macht einen „guten Cue“ aus? Strategien der Informationssuche beim heuristischen Entscheiden unter Unsicherheit*. AKZ-Forschungsbericht 05. Zürich: Angewandte Kognitionspsychologie.
- Luce, R. D. (1956). Semiorders and a theory of utility discrimination. *Econometrica*, 24, 178–191.
- Martignon, L., & Hoffrage, U. (1999). Why does one-reason decision making work? A case study in ecological rationality. In G. Gigerenzer, P. M. Todd & the ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 119–140). New York: Oxford University Press.
- Newell, B. R. (2000). *An examination of the processes underlying implicit learning*. Unpublished doctoral dissertation, University of New South Wales.
- Newell, B. R., & Shanks, D. R. (2003). Take the best or look at the rest? Factors influencing "one-reason" decision-making. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 29, 53–65.
- Newell, B. R., Rakow, T., Weston, N. J., & Shanks, D. R. (2004). Search strategies in decision-making: the success of "success". *Journal of Behavioral Decision Making*, 17, 117–137.
- Newell, B. R., Weston, N. J., & Shanks, D. R. (2003). Empirical tests of a fast and frugal heuristic: not everyone "takes-the-best". *Organizational Behavior and Human Decision Processes*, 91, 82–96.
- Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. Cambridge, MA: Cambridge University Press.
- Peterson, C. R., & Beach, L. R. (1967). Man as an intuitive statistician. *Psychological Bulletin*, 68, 29–46.
- Shuford, E. H. J. (1961). Percentage estimation of proportion as a function of element type, exposure type, and task. *Journal of Experimental Psychology*, 61, 430–436.
- Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, 69, 99–118.
- Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, 63, 129–138.
- Strerath-Bolz, U. (Red.) (2001). *Wer wird Millionär: Das Quizbuch [Who wants to be a millionaire]*. Schweizer Ausgabe. München: Knaur.
- Todd, P. M., & Gigerenzer, G. (1999). What we have learned (so far). In G. Gigerenzer, P. M. Todd & the ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 357–365). New York: Oxford University.

Teil III - Äffchen-Szenario

Kapitel 7

Das „Äffchen-Szenario“: Technische Spezifikation einer Experimentalumgebung zur Untersuchung der Lernbarkeit von Validität, Diskriminationsrate und Success von Informationsquellen

Die Güte probabilistischer Informationsquellen (in Modellen operationalisiert als Validität und Diskriminationsrate) muss in Alltagssituationen in einer konkreten Umwelt durch praktische Erfahrung gelernt werden.

Die hier beschriebene Experimentalumgebung zum Cue-Probability-Learning („Äffchen-Szenario“) erlaubt uns die Untersuchung erfahrungsabhängiger Lernvorgänge für Validitäten und Diskriminationsraten innerhalb eines Beobachtungssettings in dem V und D simultan zu erlernen sind. Spätere Aufgaben erfordern von den Personen, Tests ganz spezifisch aufgrund ihrer Validität, Diskriminationsrate oder Success-Rate (einer Kombination aus V und D) einzusetzen.

In AKZ-Forschungsbericht 37 haben wir ausführlich dargelegt, welche Überlegungen zur Entwicklung dieses Szenarios geführt haben. Im hier vorliegenden Forschungsbericht beschreiben wir nun detailliert, wie die zahlreichen Anforderungen an das Experimentaldesign auf technischer Ebene umgesetzt wurden und welche Überlegungen bei der Entwicklung der für das Experiment benötigten Bildschirm-Oberflächen massgebend waren.

7.1 Einleitung

Im Gegensatz zur TTB-Heuristik, die für die Informationssuche im Gedächtnis konzipiert wurde und nur auf den Validitäten (V) von Informationsquellen (Cues) basiert, müssen bei der (limitierten) Informationssuche in der Umwelt nicht nur die Validitäten der Cues sondern auch deren Diskriminationsraten (D) beachtet werden. Die Güte eines Cues hängt damit nicht mehr alleine von V sondern gleichzeitig auch von D ab. Dementsprechend wurden mit Usefulness ($= V \cdot D$) und Success ($= V \cdot D + [1 - D] \cdot \text{Ratewahrscheinlichkeit}$) zwei neue, aus V und D kombinierte Masse zur Berechnung der Cuegüte bei limitierter Informationssuche vorgeschlagen. Usefulness gibt an, wie häufig ein Cue eine korrekte Vorhersage macht und Success ist die geeignete Norm zur Bestimmung der Cuegüte, wenn man sich nur genau einen Cue anschauen will oder kann. Bei TTB (als gedächtnisbasierte Heuristik) muss also lediglich V durch Beobachtung möglichst genau eingeschätzt bzw. möglichst gut gelernt werden. Dem gegenüber erfordert die limitierte Informationssuche, dass entweder direkt die aus V und D kombinierten Gütemasse gelernt werden oder dass sowohl V als auch D separat gelernt und $V \cdot D$ bzw. SUC anschliessend daraus inferiert werden.

In Experimenten zur Untersuchung entscheidungspsychologischer Fragestellungen im Zusammenhang mit Informationssuche wurde und wird häufig ein Informationboard eingesetzt, auf dem Informationen abgerufen werden können. In einigen dieser Experimente wurde der Anwendungsphase eine Lernphase vorangestellt, in der nur die Validitäten gelernt werden mussten (z.B. Bröder, 2000; Newell & Shanks, 2003; Lee & Cummins, 2004; Hausmann, Christen & Läge, 2005). In mindestens einem Experiment musste auch D und Success gelernt werden (Newell, Rakow, Weston & Shanks, 2003). Bisher stand das Lernen allerdings nie im Vordergrund der Fragestellung (ausser bei Hausmann, Christen & Läge, 2005). So werden in Experimenten mit einer Lernphase die tatsächlichen Validitäten zum Zweck der experimentellen Abkürzung oft vorgegeben. Dies kann bspw. durch sporadisches Einblenden einer Liste mit den korrekten Validitäten nach einer bestimmten Anzahl Lerndurchgänge erfolgen (z.B. Newell & Shanks, 2003, Experimente 2&3) oder dadurch, dass in jedem Durchgang für jeden Cue die relativen Häufigkeiten der bis dato korrekten und falschen Vorhersagen in Balkenform dargestellt werden (Bröder, 2000). Dadurch wird es unmöglich, zu sagen, was die Vpn tatsächlich durch Beobachtung gelernt und was sie sich aus den Vorgaben gemerkt haben. Allein schon aufgrund der jeweiligen Experimentalanordnung, lässt sich mit vielen der bisher durchgeführten Experimente kein klarer Nachweis erbringen, dass Validitätsfolgen gut gelernt werden können. So konnten bspw. auch Lee & Cummins (2004) in ihrem Experiment – das ohne die Offenlegung der Validitäten auskommt – nicht zeigen, dass während einer mehrteiligen Lernphase (in der die Cue-Informationen offen vorlagen) ein Validitätslernen stattfindet. Die Autoren folgern, basierend auf der zunehmenden Anzahl der korrekten Entscheidungen (zuletzt 81%) in den verschiedenen Teilen der Lernphase, dass zumindest einige Vpn die Validitäten schnell gelernt hätten. Die untersuchten Strategien (TTB und WADD) sind jedoch schlecht gegeneinander abgegrenzt und zudem hätte bereits einfaches Tallying zu einer Rate von 83% korrekter Entscheidungen geführt. Die Schlussfolgerung, die Vpn hätten die Validitäten gelernt, ist deshalb nicht zulässig.

Beim Experiment von Hausmann, Christen & Läge (2005) steht das Lernen der Validitäten stärker im Fokus der Untersuchung. Im Gegensatz zum Experiment von Newell & Shanks (2003) liegen die Validitäten weiter auseinander und die Ergebnisse fallen hinsichtlich der Fähigkeit des Cue-Lernens deutlich positiver aus – und dies obwohl es keine direkten Hinweise auf die exakten Cue-Validitäten gibt. Die in die Lernphase eingestreuten Validitäts-Ratings (nach jeweils 10 Durchgängen) sind allerdings kritisch zu bewerten, da nicht klar ist, ob sich die Vpn bei ihrer Beurteilung jeweils auf alle Durchgänge beziehen oder nur auf die letzten 10. Die Schlussfolgerung, dass die Validitätsordnung im Wesentlichen gut gelernt wurde, kann somit nur unter Vorbehalt gezogen werden und bedarf weiterer experimenteller Bestätigung.

Newell, Rakow, Weston & Shanks (2003) argumentieren, dass das Lernen von Cügüte im Alltag gar nicht zum Wissen über Validitäten führt, sondern gleich zum integrierten Konzept des Success. Dies aufgrund der Tatsache, dass die Cügüte stets mit der Diskriminationsrate konfundiert ist. In ihren beiden Experimenten variierten sie deshalb zusätzlich zur Validität auch die Diskriminationsrate der Cues und schalteten den Anwendungsphasen jeweils eine Lernphase vor. In dieser wurden vier Cues mit ihren binären Ausprägungen gleichzeitig präsentiert. Ausserdem wurde den Vpn – wie bereits bei Newell & Shanks (2003) – nach jedem Tipp die exakte bayesianisch berechnete Wahrscheinlichkeit angegeben, mit der die gewählte Alternative die bessere sei. Die von den Autoren propagierte Alltagsnähe wird dadurch allerdings wieder fraglich, da im Alltag i.d.R. nur der tatsächliche Ausgang zusammen mit den vorliegenden Cues beobachtet werden kann, jedoch nicht die exakten, aufgrund der Validitäten bereits perfekt berechneten Wahrscheinlichkeiten. Die Auswertung des ersten Experiments zeigt, dass die gemittelte Suchreihenfolge mit der Success-Reihenfolge übereinstimmt und auch die Nachbefragung ergibt auf Stichprobenebene eine Success-Reihenfolge. 40% der Versuchspersonen haben aber offenbar nicht gelernt, welcher Cue der beste ist, da sie einen Cue bevorzugen, der nach keiner der drei Strategien (V, D, SUC) der beste ist. Aus den berichteten individuellen Daten kann nicht geschlossen werden, dass die Cue-Qualitäten systematisch gelernt wurden. Im zweiten durchgeführten Experiment wurde den Versuchspersonen nachträglich V, D und Success erklärt. Daraufhin mussten sie eine Rangordnung der vier Cues gemäss V, D und Success bilden. Aufgrund der Resultate muss geschlossen werden, dass sie V nicht gelernt hatten, dafür aber das integrierte Mass Success. Insgesamt deuten die berichteten Daten nicht darauf hin, dass V und D von den Versuchspersonen gut gelernt wird. Immerhin wird Success deutlich häufiger angewendet und auch akkurater eingeschätzt als die Validität.

Bezüglich der Lernbarkeit der Validitätsordnung, fallen die Resultate in den verschiedenen Experimenten sehr unterschiedlich aus: Sie variieren zwischen „gut“ (Hausmann et al.) und „katastrophal“ (Newell et al.). Dort wo Validitäten numerische eingeschätzt werden mussten, wurden diese substantiell unterschätzt. Auch bezüglich Success kann von einem mehrheitlichen Lernerfolg keine Rede sein. Zwar scheint die Successordnung aus dem Beobachten heraus leichter lernbar zu sein als die Validitätsordnung, auf Personenebene fallen die Resultate jedoch nicht überzeugend aus. Die Diskriminationsrate ist (in der einzigen vorhandenen Nachbefragung) zwar recht gut repräsentiert, jedoch spielt dies keine Rolle, da Success – wenn überhaupt – bereits von vornherein als eigenständiges Gütemass gelernt worden sein muss

Die bisher durchgeführten Experimente vermitteln den Eindruck, dass die Ordnungen der verschiedenen Gütemasse gar nicht oder nur unzureichend gelernt werden können. Die Experimente scheinen jedoch zu wenig auf das Erlernen der Cuegüte hin fokussiert zu sein. Deswegen legen wir hiermit ein Experiment vor, dessen Zweck einzig und allein darin besteht, die Frage nach der Lernbarkeit der Cuegüte zu klären und in seiner Konstruktion keine Kompromisse an andere, gleichzeitig zu beantwortende Fragestellungen macht.

Das Experiment besteht im Wesentlichen aus einer Lern- und einer Anwendungsphase, sowie einer Befragung, in der die Versuchspersonen die Güte der Cues bezüglich V, D und Usefulness beurteilen müssen. Die Lernsituation ist so gestaltet, dass sowohl Validität als auch Diskriminationsrate als auch Success (bzw. Usefulness) durch aktives Beobachten erlernbar sind. Die Versuchspersonen erhalten keine Hinweise auf eine bestimmte Cue-Ordnung und auch keine vorausgehende Erklärung zu den Cuegüte-Konzepten, auf die sie achten müssten. Wir weisen sie aber darauf hin, dass sie in der Anwendungsphase die Güte der Cues dringend benötigen werden. Dort geben wir ihnen die Möglichkeit, auf Verhaltensebene zu demonstrieren, ob sie Validitäten, Diskriminationsraten oder Success gelernt haben. Dabei vermeiden wir, eine Auswahl zwischen möglicherweise gelernten Cue-Rangordnungen zu erzwingen.

Auf die Frage, wie „useful“ sie die Cues fanden, gaben die Versuchspersonen im Experiment von Newell et al. Werte zwischen .20 und .30 an und tendieren damit numerisch in Richtung Usefulness (und nicht Success, da dort die Werte grundsätzlich über .50 liegen müssten). Zwar ist Success das normativ entscheidende kombinierte Mass aus Validität und Diskriminationsrate, Usefulness jedoch leichter als Konzept zu erfragen. Aus diesem Grund wählen wir genau wie Newell et al. Success als zu überprüfende Verhaltensnorm und Usefulness als Konzept der expliziten Befragung nach einem kombinierten Mass aus Validität und Diskriminationsrate.

Bei der Konstruktion der Cue-Qualitäten haben wir darauf geachtet, dass in der Verhaltensnorm Success und Usefulness auf denselben besten Cue hinweisen. Vom am höchsten diskriminierenden Cue unterscheidet sich der beste Success-Cue – anders als bei Newell et al. – hingegen deutlich. Auf diese Weise können wir – sowohl auf Verhaltensebene als auch auf der Ebene deklarativen Wissens – für jede einzelne Versuchsperson bestimmen welche der drei Gütemasse sie gelernt hat.

Auf der Basis der Argumentation von Newell et al. wäre zu erwarten, dass eine substantielle Anzahl von Personen zwar Success beherrscht, jedoch nicht sowohl Validität als auch Diskriminationsrate. Diese Personen hätten dann ein Gefühl für Success erworben, welches sich nicht aus den separat gelernten Komponenten zusammensetzt, sondern als solches durch Beobachtung gelernt wird. Sollte die TTB-Suchregel (absteigende Ordnung der Validitäten) Bestand haben, so müsste die Validitätsordnung von den Versuchspersonen gut zu lernen sein (unabhängig davon, was mit den anderen Gütemassen geschieht, denn die sind ja für TTB nicht relevant). Für nicht-kompensatorische Urteilsheuristiken reicht es aus, wenn die lexikographische Rangordnung der Cuegüte gelernt wird. Für kompensatorische, gewichtete Informationsintegration müssen die Versuchspersonen hingegen zwingend ein Gefühl für die absoluten Validitäten entwickeln.

7.2 Aufbau des Experiments

Aus technischer Sicht besteht das Experiment aus einer Ansammlung von Modulen die je ihre eigene (grafische) Benutzeroberfläche realisieren. Die folgenden Module (siehe Tabelle 7.1) wurden implementiert: Login (LG), Text (TX), Trainingsphase (TR), Anwendungsphase (AP), Präferenzen (PR), Nachbefragung (NB) und schliesslich noch das Modul VpDaten (VD) das der Erhebung persönlicher Angaben dient. Die Module können grundsätzlich allesamt mehrfach und in beliebiger Reihenfolge aufgerufen werden. Die Reihenfolge der Aufrufe ist in einer Datenbank abgespeichert und gilt für sämtliche Versuchspersonen gleichermassen. Beim Aufruf eines Moduls können verschiedene Parameter übergeben werden. Diese sind ebenfalls in der Datenbank abgelegt. So muss bspw. dem Modul TX der Pfad zu einem HTML-Dokument übergeben werden, dessen Inhalt dargestellt werden soll, und das Modul TR verlangt als Parameter u.a. die minimale und die maximale Anzahl durchzuführender Trainingsrunden.

7.3 Ablauf

Tabelle 7.2 gibt einen Überblick über die einzelnen Schritte des Experiments, die nachfolgend detailliert beschrieben werden. (Die in der 1. Spalte angegebenen Schrittnummern entsprechen gleichzeitig auch den Nummern der Abbildungen im vorliegenden Forschungsbericht.)

Tabelle 7.1: Module

Nr	Modul-Bezeichnung	ID
1	Login	LG
2	Text	TX
3	Trainingsphase	TR
4	Anwendungsphase	AP
5	Präferenzen	PR
6	Nachbefragung	NB
7	VpDaten	VD

Tabelle 7.2: Schritte des Experiments

Schritt	Mo-	Beschreibung
01	LG	Login
02	TX	Begrüssung
03	TX	Spielanleitung Seite 1
04	TX	Spielanleitung Seite 2
05	TX	Spielanleitung Seite 3
06	TX	Trainingsphase - Start (Info)
07	TR	Trainingsphase
08	TX	Trainingsphase - Ende (Info)
09	TX	Anwendungsphase - Start (Info)
10	TX	Expedition 1 - Start (Info)
11	AP	Expedition 1
12	TX	Expedition 1 - Ende (Info)
13	TX	Expedition 2 - Start (Info)
14	AP	Expedition 2
15	TX	Expedition 2 - Ende (Info)
16	TX	Expedition 3 - Start (Info)
17	AP	Expedition 3
18	TX	Expedition 3 - Ende (Info)
19	TX	Präferenzen - Start (Info)
20	PR	Präferenzen
21	TX	Nachbefragung - Start (Info)
22	NB	Nachbefragung
23	TX	Persönliche Daten - Start (Info)
24	VD	Persönliche Daten
25	TX	Schlussseite (Info)
26	LG	Login

Login: Zum Login gibt es nicht viel zu sagen. Der Versuchsleiter wählt seinen Namen aus dem drop-down Bereich des oberen Listenfelds aus und gibt den Namen der Versuchsperson (oder eine anonymisierte Bezeichnung derselben) in das zweite Listenfeld ein (siehe Abbildung 7.1).

Abbildung 7.1: Login

Begrüssung (Willkommenseite): Auf das Login folgt die Anzeige eines Begrüssungstexts. Dort bedanken wir uns im Voraus für die Bereitschaft zur Teilnahme am Experiment und kündigen die auf den nachfolgenden Seiten erscheinende Spielanleitung an (siehe Abbildung 7.2).

Abbildung 7.2: Begrüssungstext

Anleitung zum Experiment (Rahmengeschichte): Die Anleitung umfasst drei Bildschirmseiten, zwischen denen beliebig oft vor und zurück geschaltet werden kann. Sie präsentiert die Rahmengeschichte und weist auf die Lotterie³⁴ hin, an der das beste Drittel der Versuchspersonen nach

³⁴ Gehörte die Versuchsperson zum Drittel derjenigen Probanden, die über die Summe der 3 Expeditionen die meisten Äffchen gerettet hatten, so nahm sie an einer Verlosung von 3 x 100 Schweizer Franken teil. Den Versuchspersonen wurde dieses am Ende der Instruktion mitgeteilt, so dass eine moderate Leistungsmotivation induziert war.

Abschluss der Datenerhebung automatisch teilnimmt. (siehe Abbildungen 7.3 bis 7.5 sowie Anhänge, 7.8.1: Rahmengeschichte). Die Anleitung kann während des Experiments jederzeit wieder eingeblendet werden – ein entsprechender Hinweis findet sich am Ende der Anleitung (vgl. Abbildung 7.5). Sie soll die Versuchspersonen möglichst gut in die Experimental-Situation einführen und auf die durchzuführenden Aufgaben vorbereiten. Die Datenerhebung besteht im Wesentlichen aus zwei Phasen, einer Lernphase (Trainingsphase) und einer aus drei Teilen bestehenden Anwendungsphase (Expeditionsphase). In beiden Phasen steht den Versuchspersonen ein Information-Board mit fünf binären Cues (medizinische Tests) zur Informationssuche zur Verfügung. Mit Hilfe der Tests sollen die Versuchsperson den Gesundheitszustand von Äffchen beurteilen und eine Behandlungsentscheidung (Behandlung vs. keine Behandlung) treffen, wobei bekannt ist, dass ca. 50% der Äffchen unter einer tödlich verlaufenden Krankheit leiden und die anderen 50% vollständig gesund sind. Kurzum, es geht darum, erkrankte Äffchen zu identifizieren und (mittels einmaliger Verabreichung eines bestimmten Medikaments) zu behandeln. Dabei gilt es zu beachten, dass das Medikament möglichst nur tatsächlich erkrankten Äffchen appliziert wird, da gesunde Äffchen mit Sicherheit sterben, sofern sie mit dem Medikament behandelt werden³⁵.

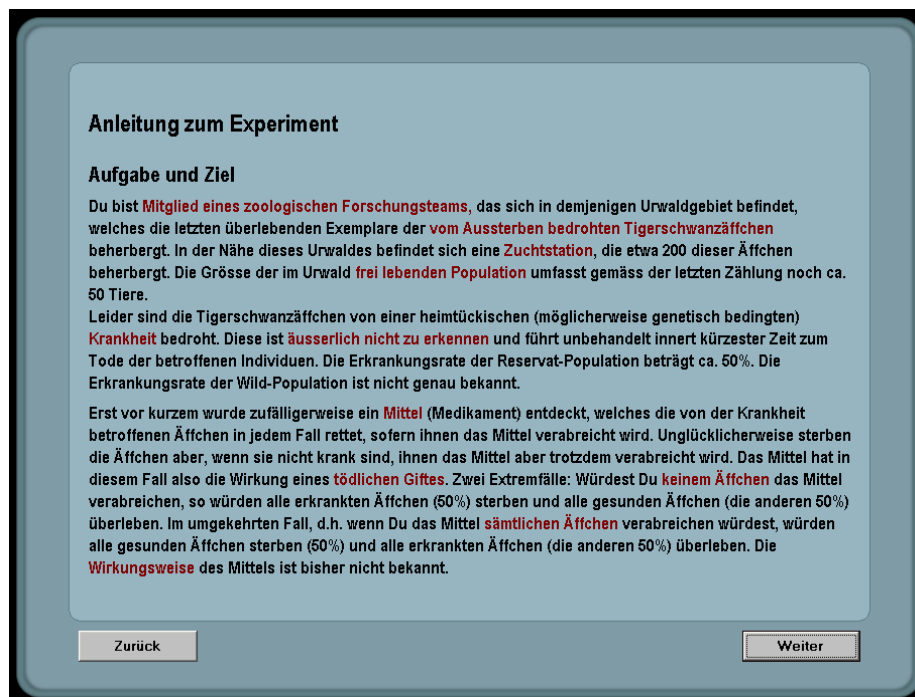


Abbildung 7.3: Spielanleitung (Seite 1)

³⁵ Im Zusammenspiel mit der auf 50% festgelegten Erkrankungsrate erreichen wir durch die Einführung dieser Rahmenbedingung, dass sich eine reine Ratestrategie oder eine Extremstrategie (wie bspw. die, allen Äffchen das Medikament unbesehen zu verabreichen), nicht lohnt. Würde eine Versuchsperson jedes Mal raten, ob sie ein Äffchen behandeln soll oder nicht, läge ihre mittlere Erfolgsquote bei 50%. Würde durchgehend jedes oder durchgehend keines der Äffchen behandelt, läge die Erfolgsquote ebenfalls bei 50% – diesmal allerdings nicht nur im Mittel sondern konstant.

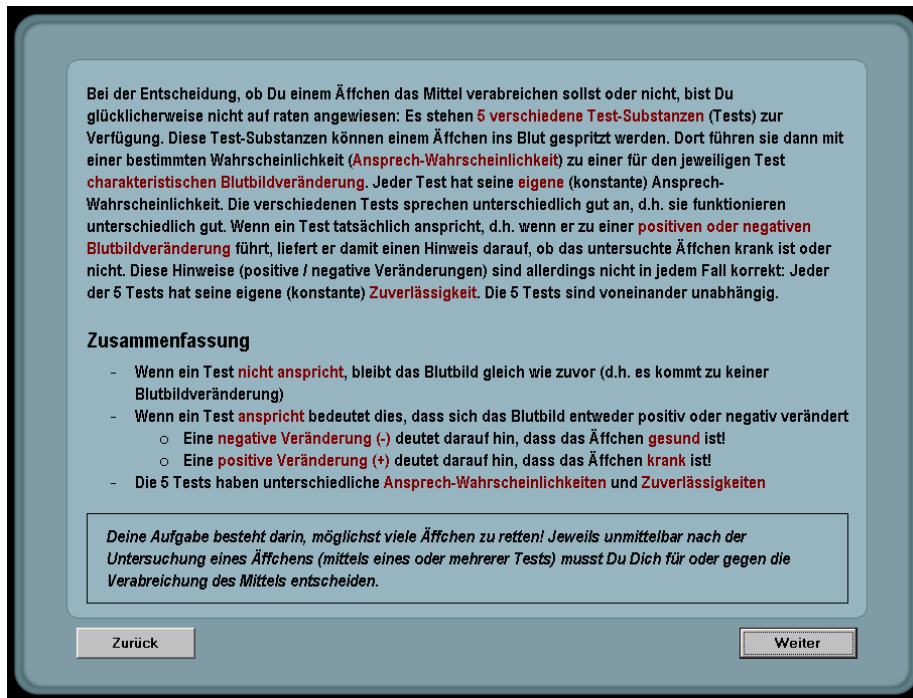


Abbildung 7.4: Spielanleitung (Seite 2)

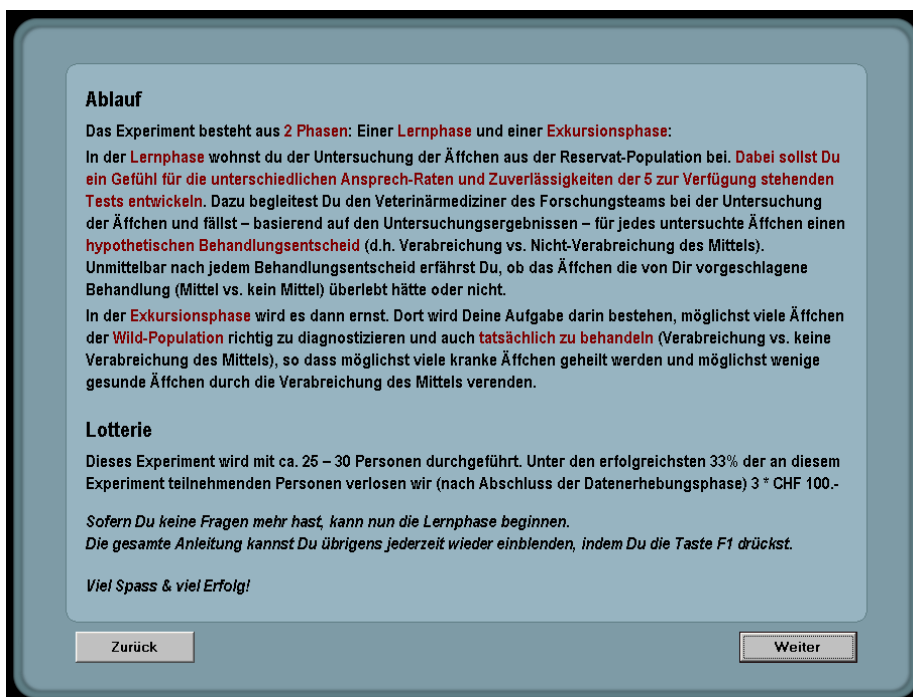


Abbildung 7.5: Spielanleitung (Seite 3)

Hinweis zu Beginn der Lernphase (Trainingsphase): Vor dem Start der Lernphase weisen wir die Versuchspersonen implizit auf die Bedeutung der Zuverlässigkeiten (Validitäten) und der Ansprechraten (Diskriminationsraten) der medizinischen Tests hin, indem wir sie auffordern diese während der Lernphase zu beachten, da sie später (in der Anwendungsphase) in der Regel nicht

mehr alle 5 Tests werden einsetzen können bzw. nicht mehr alle Cue-Informationen automatisch vorliegen (siehe Abbildung 7.6).



Abbildung 7.6: Trainingsphase (Start)

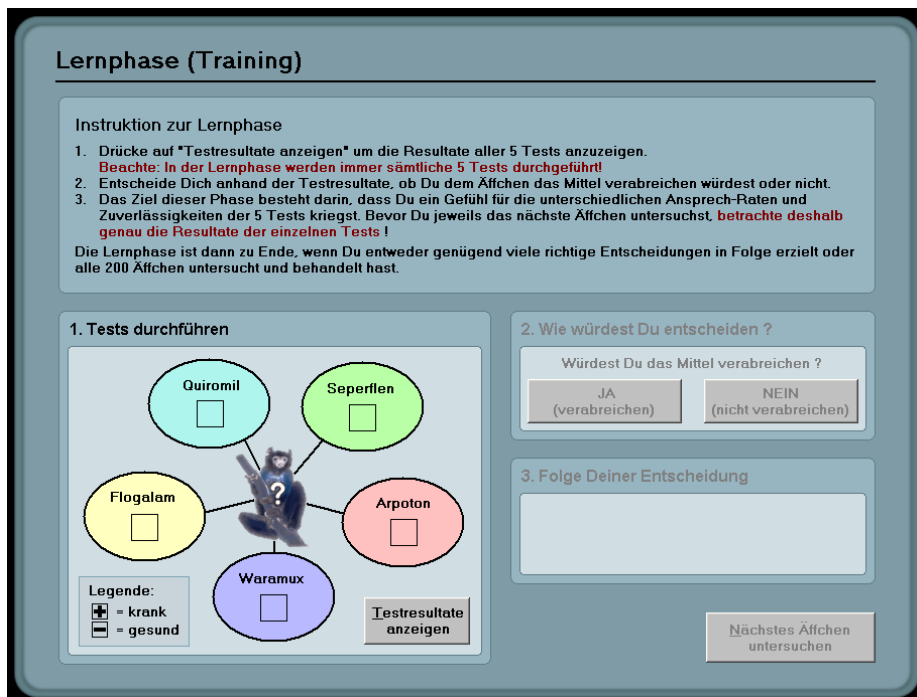


Abbildung 7.7a: Trainingsphase. Auf der linken Seite, unterhalb der Instruktion sind die fünf Cues (Tests) zu sehen. Zur Anzeige der Testresultate muss die Schaltfläche "Testresultate anzeigen" gedrückt werden.

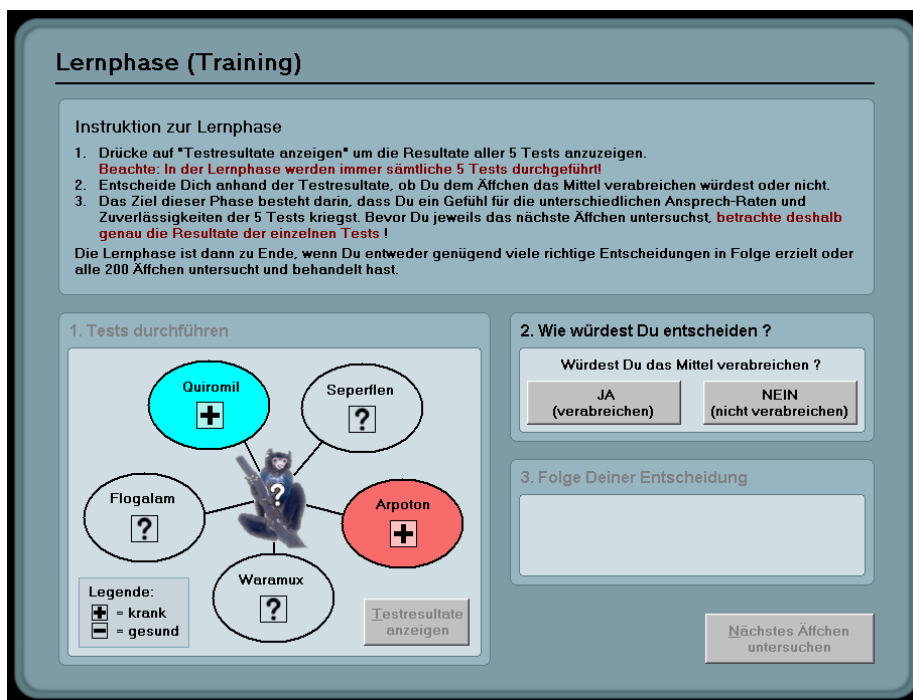


Abbildung 7.7b: Trainingsphase (Beispiel). Nachdem die Schaltfläche „Testresultate anzeigen“ gedrückt wurde, muss sich die Versuchsperson für oder gegen die Verabreichung des Mittels entscheiden

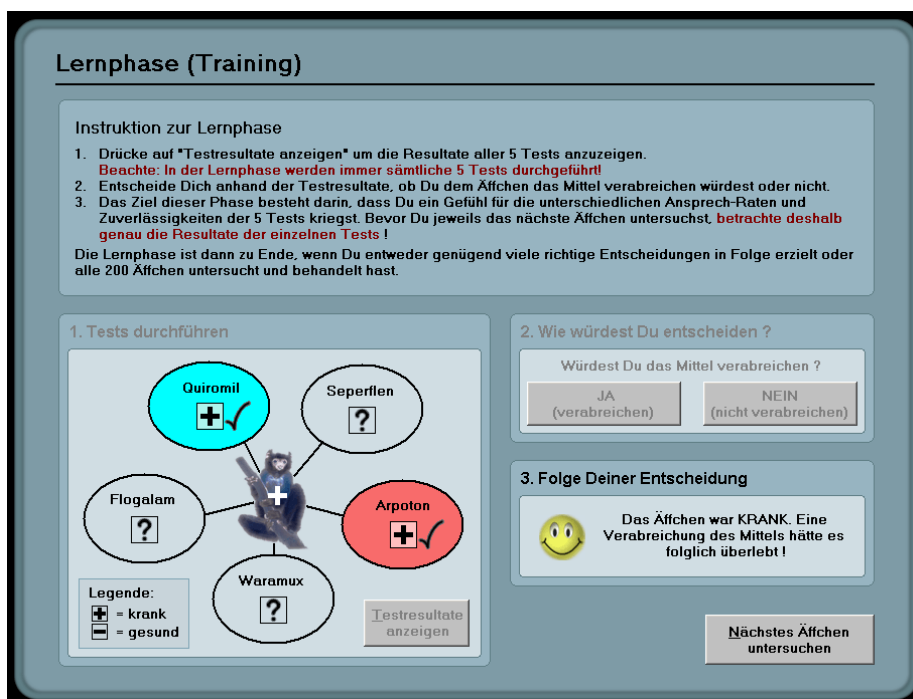


Abbildung 7.7c: Trainingsphase (Beispiel). Nach der Entscheidung für oder gegen die Verabreichung des Mittels, wird die Folge der Entscheidung mitgeteilt (rechte Seite, unten) und gleichzeitig für jeden diskriminierenden Test angezeigt, ob sein Hinweis korrekt (Häkchen) oder falsch (Kreuz) war.

Lernphase (Trainingsphase): In der Lernphase können die Versuchspersonen die Arbeit von Veterinärmedizinern an einer in einem grossen Freigehege gehaltenen Äffchen-Population beobachten. Für diese Population ist bekannt, dass ca. 50% der Äffchen unter der Krankheit leiden. Die Versuchspersonen erhalten in jeder Runde (also für jedes getestete Äffchen) immer alle Cue-Informationen (Test-Resultate) auf dem Bildschirm angezeigt, wobei ein „+“ für „Das Äffchen ist krank“ und ein „-“ für „Das Äffchen ist gesund“ steht. In der Regel sprechen aber nicht immer alle Tests gleichzeitig an. Falls ein Test nicht anspricht, wird anstelle eines Test-Resultats lediglich ein Fragezeichen („?“) angezeigt. Die Versuchspersonen können dann einen Tipp abgeben, ob das Äffchen erkrankt sei oder nicht. Ihnen wird jedoch gesagt, dass dieser Tipp reine Übungszwecke habe und keinen Einfluss darauf besitze, ob das riskante Medikament appliziert werde oder nicht. Als Ergebnis erhalten sie rückgemeldet, ob das Äffchen tatsächlich krank ist oder nicht. Dieses Ergebnis kann mit dem eigenen Tipp und vor allem mit den fünf Testresultaten verglichen werden. Dann geht es weiter zur nächsten Aufgabe. So erhalten die Versuchspersonen die Möglichkeit, eine Repräsentation der über das gesamte Experiment konstant gehaltenen Validitäten und Diskriminationsraten der fünf Cues aufzubauen (siehe Abbildungen 7.7a–7.7c).

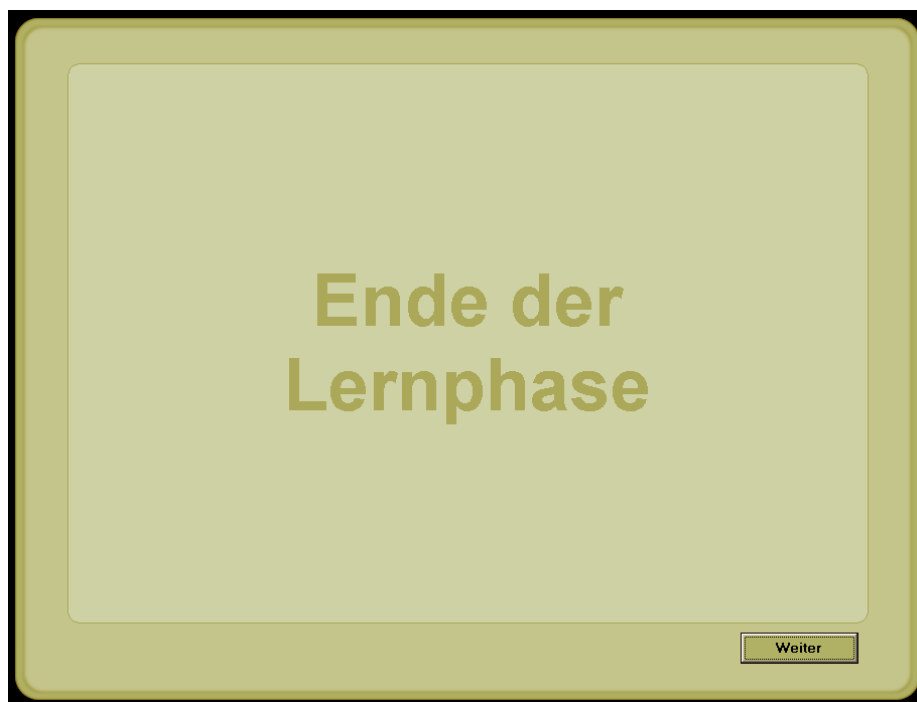


Abbildung 7.8: Das Ende der Lernphase wird angezeigt.

Ende der Trainingsphase (Abbildung 7.8): Die Trainingsphase umfasste – abhängig von der Lerngeschwindigkeit einer Versuchsperson – mindestens 100 jedoch maximal 200 Durchgänge. Das Ende ist erreicht, sobald eine Versuchsperson insgesamt 100 richtige Entscheidungen getroffen oder 200 Trainingsrunden absolviert hat.

Hinweis zur Anwendungsphase: Auf dieser Seite künden wir den Beginn der Anwendungsphase an, repetieren kurz um was es geht und weisen darauf hin, dass die Tests ab nun nur noch einzelnen durchgeführt werden können (siehe Abbildung 7.9).



Abbildung 7.9: Erläuterungen vor Beginn der Expeditionsphase („Exkursionsphase“).

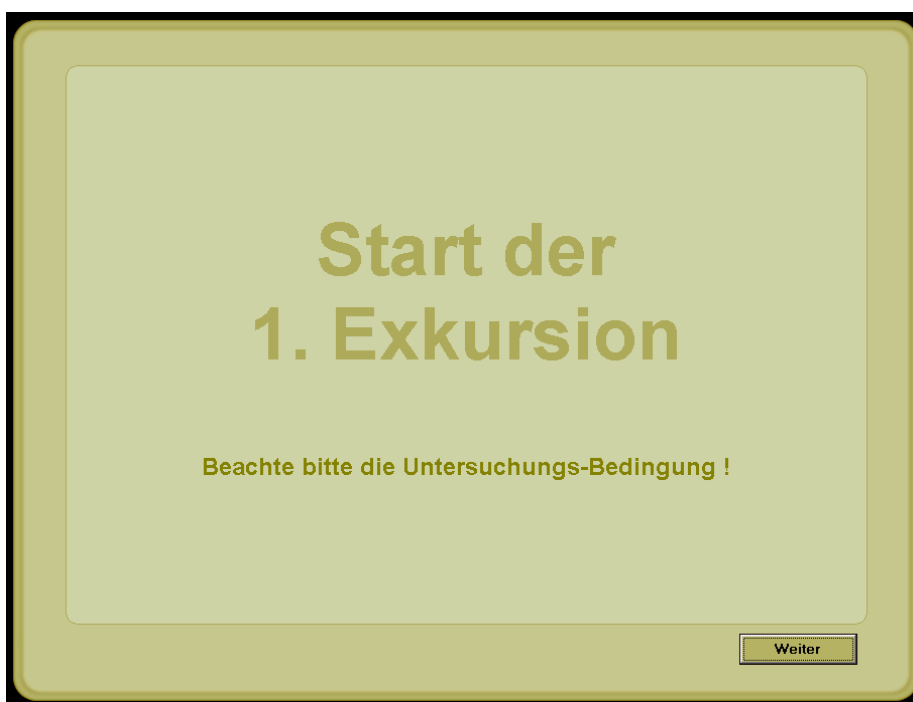


Abbildung 7.10: Beginn von Expedition 1 („Exkursion 1“).

1. Exkursion

Untersuchungs-Bedingung für die 1. Exkursion

Gemäss Anweisungen der Organisation für die Du arbeitest, musst Du die **Untersuchung eines Äffchens** aus finanziellen Gründen **sofort abbrechen** und einen Behandlungsentscheid treffen, **sobald ein Test an diesem Äffchen angesprochen hat** (d.h. zu einer positiven oder negativen Blutbildveränderung geführt hat). Deine Entscheidung, ein Äffchen zu behandeln oder nicht zu behandeln, basiert somit auf jeweils genau einem einzigen Testergebnis. Die Anzahl an einem Äffchen durchführbarer Tests ist **nicht grundsätzlich beschränkt**. Du kannst jedoch **keine weiteren Tests mehr an dem betreffenden Äffchen** durchführen, nachdem Du ein erstes (positives oder negatives) Testergebnis für dieses Äffchen erhalten hast.

Wähle deshalb die Reihenfolge der Tests gut aus !

1. Tests durchführen

Legende:

☒ = krank

☐ = gesund

2. Wie entscheidest Du Dich ?

Wirst Du das Mittel verabreichen ?

3. Folge Deiner Entscheidung

Abbildung 7.11a: Expedition 1 („Exkursion 1“). Noch wurden keine Tests durchgeführt

1. Exkursion

Untersuchungs-Bedingung für die 1. Exkursion

Gemäss Anweisungen der Organisation für die Du arbeitest, musst Du die **Untersuchung eines Äffchens** aus finanziellen Gründen **sofort abbrechen** und einen Behandlungsentscheid treffen, **sobald ein Test an diesem Äffchen angesprochen hat** (d.h. zu einer positiven oder negativen Blutbildveränderung geführt hat). Deine Entscheidung, ein Äffchen zu behandeln oder nicht zu behandeln, basiert somit auf jeweils genau einem einzigen Testergebnis. Die Anzahl an einem Äffchen durchführbarer Tests ist **nicht grundsätzlich beschränkt**. Du kannst jedoch **keine weiteren Tests mehr an dem betreffenden Äffchen** durchführen, nachdem Du ein erstes (positives oder negatives) Testergebnis für dieses Äffchen erhalten hast.

Wähle deshalb die Reihenfolge der Tests gut aus !

1. Tests durchführen

Legende:

☒ = krank

☐ = gesund

2. Wie entscheidest Du Dich ?

Wirst Du das Mittel verabreichen ?

3. Folge Deiner Entscheidung

Abbildung 7.11b: Beispiel aus Expedition 1. Der Test „Waramux“ wurde durchgeführt, lieferte aber kein Ergebnis.

1. Exkursion

Untersuchungs-Bedingung für die 1. Exkursion

Gemäss Anweisungen der Organisation für die Du arbeitest, musst Du die **Untersuchung eines Äffchens** aus finanziellen Gründen **sofort abbrechen** und einen Behandlungsentscheid treffen, **sobald ein Test an diesem Äffchen angesprochen hat** (d.h. zu einer positiven oder negativen Blutbildveränderung geführt hat). Deine Entscheidung, ein Äffchen zu behandeln oder nicht zu behandeln, basiert somit auf jeweils genau einem einzigen Testergebnis. Die Anzahl an einem Äffchen durchführbarer Tests ist **nicht grundsätzlich beschränkt**. Du kannst jedoch **keine weiteren Tests mehr an dem betreffenden Äffchen** durchführen, nachdem Du ein erstes (positives oder negatives) Testergebnis für dieses Äffchen erhalten hast.
Wähle deshalb die Reihenfolge der Tests gut aus !

1. Tests durchführen

Legende:
 + = krank
 - = gesund

2. Wie entscheidest Du Dich ?

Wirst Du das Mittel verabreichen ?

JA
(verabreichen)

NEIN
(nicht verabreichen)

3. Folge Deiner Entscheidung

Nächstes Äffchen untersuchen

Abbildung 7.11c: Beispiel aus Expedition 1. Auch der Test „Seperflen“ lieferte kein Ergebnis.

1. Exkursion

Untersuchungs-Bedingung für die 1. Exkursion

Gemäss Anweisungen der Organisation für die Du arbeitest, musst Du die **Untersuchung eines Äffchens** aus finanziellen Gründen **sofort abbrechen** und einen Behandlungsentscheid treffen, **sobald ein Test an diesem Äffchen angesprochen hat** (d.h. zu einer positiven oder negativen Blutbildveränderung geführt hat). Deine Entscheidung, ein Äffchen zu behandeln oder nicht zu behandeln, basiert somit auf jeweils genau einem einzigen Testergebnis. Die Anzahl an einem Äffchen durchführbarer Tests ist **nicht grundsätzlich beschränkt**. Du kannst jedoch **keine weiteren Tests mehr an dem betreffenden Äffchen** durchführen, nachdem Du ein erstes (positives oder negatives) Testergebnis für dieses Äffchen erhalten hast.
Wähle deshalb die Reihenfolge der Tests gut aus !

1. Tests durchführen

Legende:
 + = krank
 - = gesund

2. Wie entscheidest Du Dich ?

Wirst Du das Mittel verabreichen ?

JA
(verabreichen)

NEIN
(nicht verabreichen)

3. Folge Deiner Entscheidung

Nächstes Äffchen untersuchen

Abbildung 7.11d: Beispiel aus Expedition 1. Das Ergebnis des dritten durchgeführten Tests („Flogalam“) deutet darauf hin, dass das Äffchen gesund ist.

1. Exkursion

Untersuchungs-Bedingung für die 1. Exkursion

Gemäss Anweisungen der Organisation für die Du arbeitest, musst Du die **Untersuchung eines Äffchens** aus finanziellen Gründen **sofort abbrechen** und einen Behandlungsentscheid treffen, **sobald ein Test an diesem Äffchen angesprochen hat** (d.h. zu einer positiven oder negativen Blutbildveränderung geführt hat). Deine Entscheidung, ein Äffchen zu behandeln oder nicht zu behandeln, basiert somit auf jeweils genau einem einzigen Testergebnis. Die Anzahl an einem Äffchen durchführbarer Tests ist **nicht grundsätzlich beschränkt**. Du kannst jedoch **keine weiteren Tests mehr an dem betreffenden Äffchen** durchführen, nachdem Du ein erstes (positives oder negatives) Testergebnis für dieses Äffchen erhalten hast.
Wähle deshalb die Reihenfolge der Tests gut aus !

1. Tests durchführen

Legende:

+ = krank

- = gesund

2. Wie entscheidest Du Dich ?

Wirst Du das Mittel verabreichen ?

JA
(verabreichen)

NEIN
(nicht verabreichen)

3. Folge Deiner Entscheidung

Das Äffchen war **GESUND**. Dank Deiner Entscheidung, ihm das Mittel **NICHT** zu verabreichen, hat es überlebt !

Nächstes Äffchen untersuchen

Abbildung 7.11e: Beispiel aus Expedition 1. Die Versuchsperson entscheidet sich, dem Äffchen das Mittel nicht zu verabreichen und liegt damit richtig, da das Äffchen bereits gesund war.

Ende der 1. Exkursion

Weiter

Abbildung 7.12: Ende von Expedition 1.

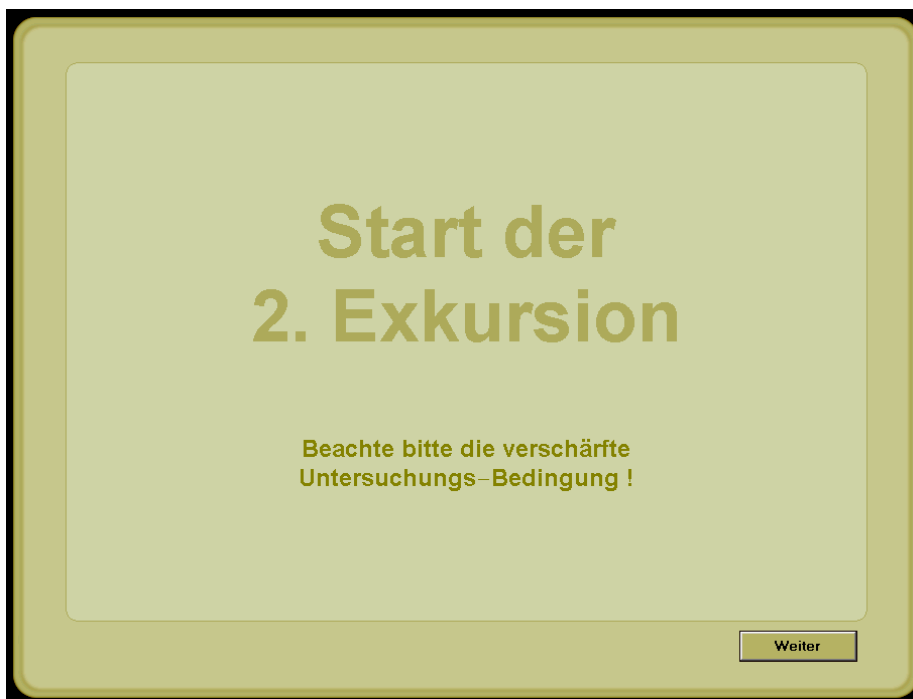


Abbildung 7.13: Beginn von Expedition 2.

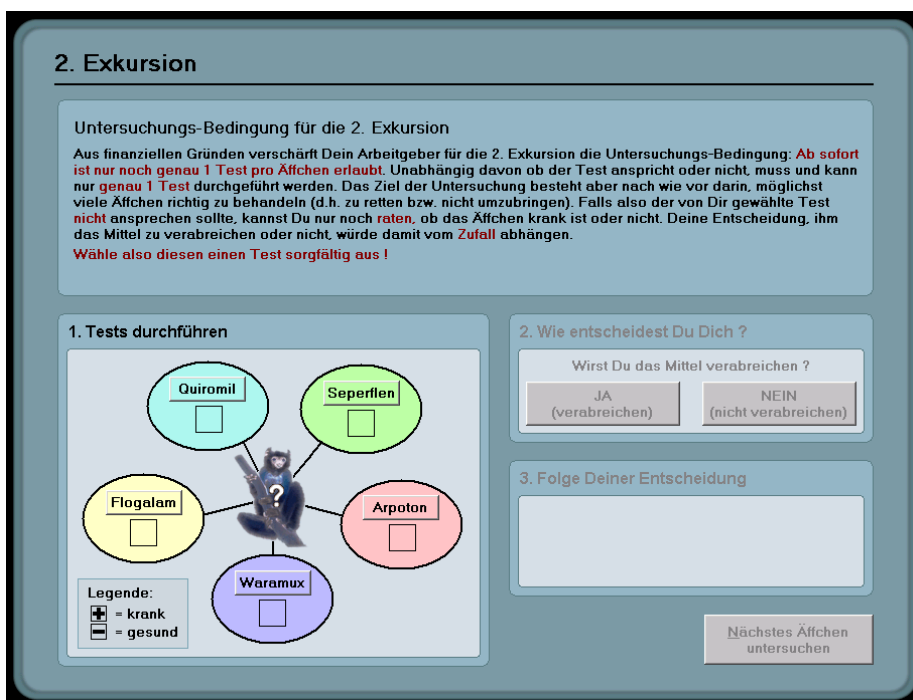


Abbildung 7.14a: Expedition 2 („Exkursion 2“). Noch wurden keine Tests durchge-

2. Exkursion

Untersuchungs-Bedingung für die 2. Exkursion

Aus finanziellen Gründen verschärft Dein Arbeitgeber für die 2. Exkursion die Untersuchungs-Bedingung: **Ab sofort ist nur noch genau 1 Test pro Äffchen erlaubt.** Unabhängig davon ob der Test anspricht oder nicht, muss und kann nur **genau 1 Test** durchgeführt werden. Das Ziel der Untersuchung besteht aber nach wie vor darin, möglichst viele Äffchen richtig zu behandeln (d.h. zu retten bzw. nicht umzubringen). Falls also der von Dir gewählte Test **nicht** ansprechen sollte, kannst Du nur noch **raten**, ob das Äffchen krank ist oder nicht. Deine Entscheidung, ihm das Mittel zu verabreichen oder nicht, würde damit vom **Zufall** abhängen.

Wähle also diesen einen Test sorgfältig aus !

1. Tests durchführen

Legende:
 + = krank
 - = gesund

2. Wie entscheidest Du Dich ?

Wirst Du das Mittel verabreichen ?

JA (verabreichen) NEIN (nicht verabreichen)

3. Folge Deiner Entscheidung

Nächstes Äffchen untersuchen

Abbildung 7.14b: Beispiel aus Expedition 2 („Exkursion 2“). Der erste durchgeführte Test hat angeschlagen und deutet darauf hin, dass das Äffchen krank ist.

2. Exkursion

Untersuchungs-Bedingung für die 2. Exkursion

Aus finanziellen Gründen verschärft Dein Arbeitgeber für die 2. Exkursion die Untersuchungs-Bedingung: **Ab sofort ist nur noch genau 1 Test pro Äffchen erlaubt.** Unabhängig davon ob der Test anspricht oder nicht, muss und kann nur **genau 1 Test** durchgeführt werden. Das Ziel der Untersuchung besteht aber nach wie vor darin, möglichst viele Äffchen richtig zu behandeln (d.h. zu retten bzw. nicht umzubringen). Falls also der von Dir gewählte Test **nicht** ansprechen sollte, kannst Du nur noch **raten**, ob das Äffchen krank ist oder nicht. Deine Entscheidung, ihm das Mittel zu verabreichen oder nicht, würde damit vom **Zufall** abhängen.

Wähle also diesen einen Test sorgfältig aus !

1. Tests durchführen

Legende:
 + = krank
 - = gesund

2. Wie entscheidest Du Dich ?

Wirst Du das Mittel verabreichen ?

JA (verabreichen) NEIN (nicht verabreichen)

3. Folge Deiner Entscheidung

Das Äffchen war GESUND. Aufgrund Deiner Entscheidung, ihm das Mittel zu verabreichen, ist es leider verstorben !

Nächstes Äffchen untersuchen

Abbildung 7.14c: Beispiel aus Expedition 2 („Exkursion 2“). Durch die Verabreichung des Mittels ist das Äffchen leider verstorben, da es bereits gesund war.

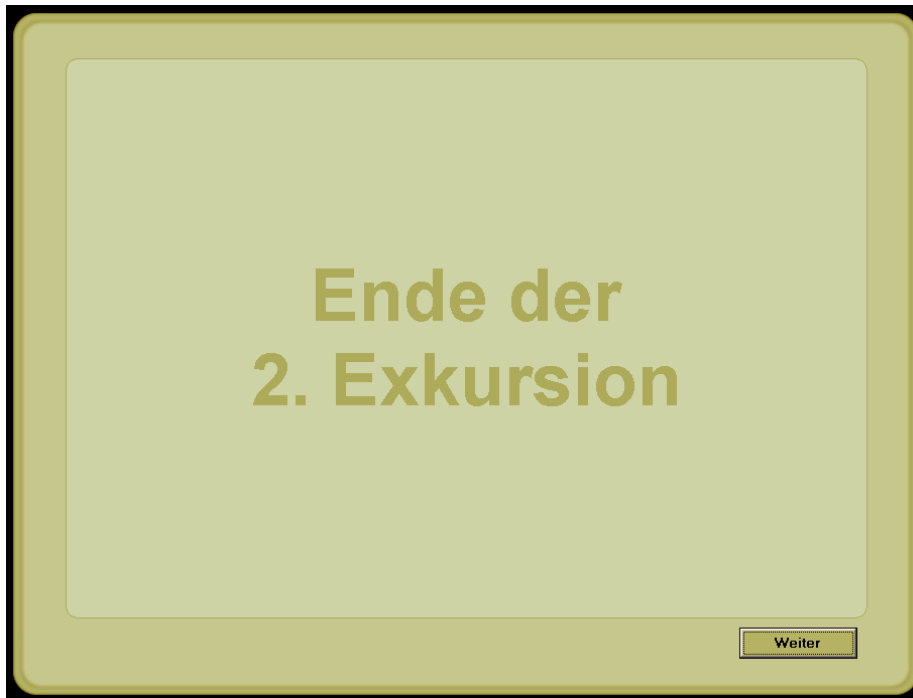


Abbildung 7.15: Ende von Expedition 2 („Exkursion 2“).

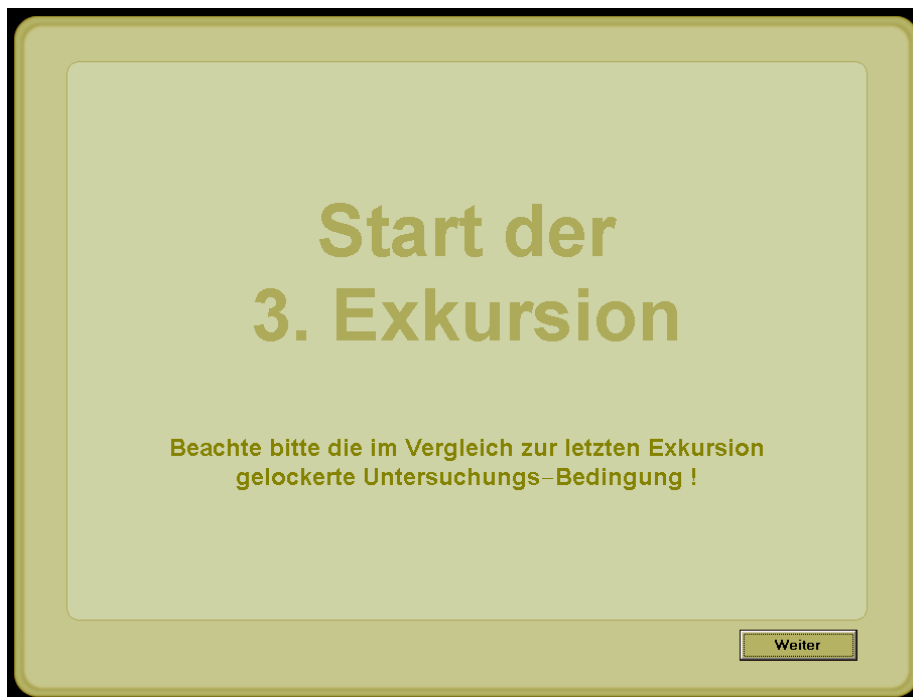


Abbildung 7.16: Beginn von Expedition 3 („Exkursion 3“).

3. Exkursion

Untersuchungs-Bedingung für die 3. Exkursion

Für die 3. Exkursion lockert Dein Arbeitgeber die Untersuchungs-Bedingung nun wieder ein wenig. Bezüglich der Anzahl erlaubter Tests pro Äffchen gilt nun folgendes: Sofern **der erste** an einem Äffchen durchgeführte Test anspricht, darfst Du an diesem Äffchen **beliebig viele weitere Tests** durchführen (da, aus rechtlichen Gründen, nur in diesem Fall zusätzliche, gebundene Spendengelder für die weiteren Tests zur Verfügung stehen). Falls jedoch der erste von Dir gewählte Test **nicht anspricht**, darfst Du **keine weiteren Tests mehr** durchführen. Das Überleben des betreffenden Äffchens würde damit von einer **Zufallsentscheidung** abhängen.

Wähle deshalb auch hier die Reihenfolge der Tests (insbesondere des ersten) sorgfältig aus !

1. Tests durchführen

Legende:
 + = krank
 - = gesund

2. Wie entscheidest Du Dich ?

Wirst Du das Mittel verabreichen ?

JA
(verabreichen)

NEIN
(nicht verabreichen)

3. Folge Deiner Entscheidung

Nächstes Äffchen untersuchen

Abbildung 7.17a: Expedition 3 („Exkursion 3“). Noch wurden keine Tests durchge-

3. Exkursion

Untersuchungs-Bedingung für die 3. Exkursion

Für die 3. Exkursion lockert Dein Arbeitgeber die Untersuchungs-Bedingung nun wieder ein wenig. Bezüglich der Anzahl erlaubter Tests pro Äffchen gilt nun folgendes: Sofern **der erste** an einem Äffchen durchgeführte Test anspricht, darfst Du an diesem Äffchen **beliebig viele weitere Tests** durchführen (da, aus rechtlichen Gründen, nur in diesem Fall zusätzliche, gebundene Spendengelder für die weiteren Tests zur Verfügung stehen). Falls jedoch der erste von Dir gewählte Test **nicht anspricht**, darfst Du **keine weiteren Tests mehr** durchführen. Das Überleben des betreffenden Äffchens würde damit von einer **Zufallsentscheidung** abhängen.

Wähle deshalb auch hier die Reihenfolge der Tests (insbesondere des ersten) sorgfältig aus !

1. Tests durchführen

Legende:
 + = krank
 - = gesund

2. Wie entscheidest Du Dich ?

Wirst Du das Mittel verabreichen ?

JA
(verabreichen)

NEIN
(nicht verabreichen)

3. Folge Deiner Entscheidung

Nächstes Äffchen untersuchen

Abbildung 7.17b: Beispiel aus Expedition 3 („Exkursion 3“). Der erste durchgeführte Test („Arpoton“) hat angeschlagen. Gemäss Untersuchungsbedingung dürfen deshalb auch die restlichen vier Tests noch durchgeführt werden.

3. Exkursion

Untersuchungs-Bedingung für die 3. Exkursion

Für die 3. Exkursion lockert Dein Arbeitgeber die Untersuchungs-Bedingung nun wieder ein wenig. Bezüglich der Anzahl erlaubter Tests pro Äffchen gilt nun folgendes: Sofern **der erste** an einem Äffchen durchgeführte Test anspricht, darfst Du an diesem Äffchen **beliebig viele weitere Tests** durchführen (da, aus rechtlichen Gründen, nur in diesem Fall zusätzliche, gebundene Spendengelder für die weiteren Tests zur Verfügung stehen). Falls jedoch der erste von Dir gewählte Test **nicht anspricht**, darfst Du **keine weiteren Tests mehr** durchführen. Das Überleben des betreffenden Äffchens würde damit von einer **Zufallsentscheidung** abhängen.

Wähle deshalb auch hier die Reihenfolge der Tests (insbesondere des ersten) sorgfältig aus !

1. Tests durchführen

Legende:
+ = krank
- = gesund

2. Wie entscheidest Du Dich ?

Wirst Du das Mittel verabreichen ?

3. Folge Deiner Entscheidung

Abbildung 7.17c: Beispiel aus Expedition 3 („Exkursion 3“). Von den restlichen vier Tests schlägt nur Quiromil an und auch dieser Test deutet darauf hin, dass das Äffchen krank ist.

3. Exkursion

Untersuchungs-Bedingung für die 3. Exkursion

Für die 3. Exkursion lockert Dein Arbeitgeber die Untersuchungs-Bedingung nun wieder ein wenig. Bezüglich der Anzahl erlaubter Tests pro Äffchen gilt nun folgendes: Sofern **der erste** an einem Äffchen durchgeführte Test anspricht, darfst Du an diesem Äffchen **beliebig viele weitere Tests** durchführen (da, aus rechtlichen Gründen, nur in diesem Fall zusätzliche, gebundene Spendengelder für die weiteren Tests zur Verfügung stehen). Falls jedoch der erste von Dir gewählte Test **nicht anspricht**, darfst Du **keine weiteren Tests mehr** durchführen. Das Überleben des betreffenden Äffchens würde damit von einer **Zufallsentscheidung** abhängen.

Wähle deshalb auch hier die Reihenfolge der Tests (insbesondere des ersten) sorgfältig aus !

1. Tests durchführen

Legende:
+ = krank
- = gesund

2. Wie entscheidest Du Dich ?

Wirst Du das Mittel verabreichen ?

3. Folge Deiner Entscheidung

Das Äffchen war KRANK. Dank Deiner Entscheidung, ihm das Mittel zu verabreichen, hat es überlebt !

Abbildung 7.17d: Beispiel aus Expedition 3 („Exkursion 3“). Die Versuchsperson entscheidet sich, dem Äffchen das Mittel zu verabreichen und rettet damit das Äffchen, da es tatsächlich krank war.

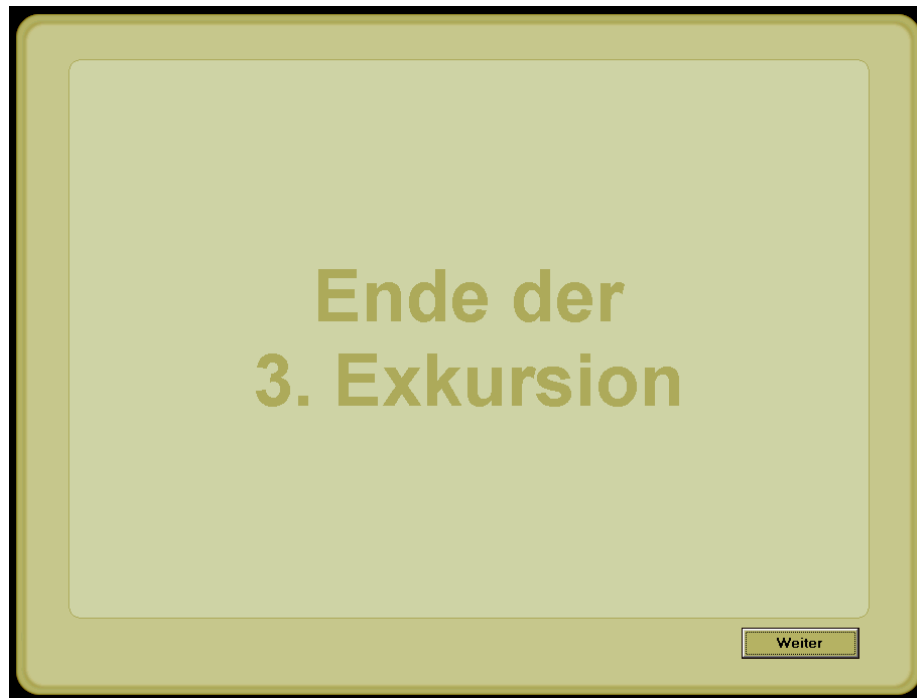


Abbildung 7.18: Ende von Expedition 3 („Exkursion 3“).

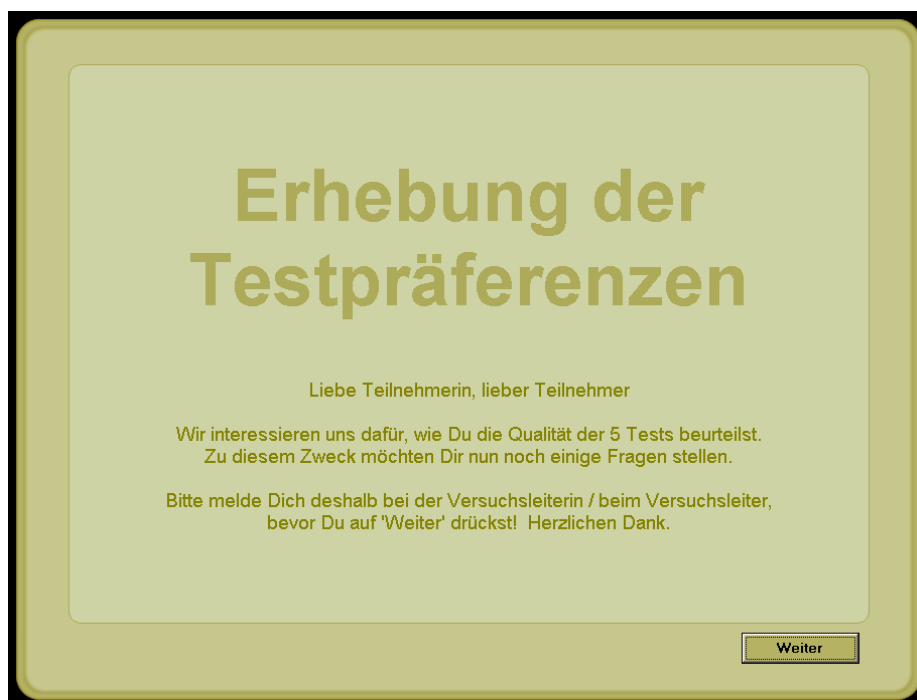


Abbildung 7.19: Hinweis auf die folgende Erhebung der Testpräferenzen

Fragen 1 2 3

1. Relative Häufigkeit korrekter Testergebnisse ?

Angenommen, Du würdest jeden einzelnen Test so oft hintereinander anwenden, bis dass er insgesamt 100-mal angesprochen hat, welcher der 5 Tests würde dann am häufigsten, zweit-häufigsten, etc. ein zuverlässiges Ergebnis liefern ?

am zuverlässigsten :

1.

2.

3.

4.

am unzuverlässigsten :

5.

O S F A W

Weiter

Abbildung 7.20a: Erhebung der Testpräferenzen, Aufgabe 1a: Die Tests sollen in eine Rangreihe gebracht werden. Das zu berücksichtigende Kriterium ist die Validität der Tests.

Fragen 1 2 3

1. Relative Häufigkeit korrekter Testergebnisse ?

Angenommen, Du würdest jeden einzelnen Test so oft hintereinander anwenden, bis dass er insgesamt 100-mal angesprochen hat, welcher der 5 Tests würde dann am häufigsten, zweit-häufigsten, etc. ein zuverlässiges Ergebnis liefern ?

am zuverlässigsten :

1.

2.

3.

4.

am unzuverlässigsten :

5.

Flogalam
Arpton
Quiromil

O S F A W

Weiter

Abbildung 7.20b: Erhebung der Testpräferenzen, Aufgabe 1a: Bildung einer Rangreihe gemäss dem Kriterium der Validität der Tests (Beispiel).

Fragen 1 2 3

1. Relative Häufigkeit korrekter Testergebnisse ?

Angenommen, Du würdest jeden einzelnen Test so oft hintereinander anwenden, bis dass er insgesamt 100-mal angesprochen hat, welcher der 5 Tests würde dann am häufigsten, zweit-häufigsten, etc. ein zuverlässiges Ergebnis liefern ?

am zuverlässigsten :

1. Waramux

2. Seperflen

3. Flogalam

4. Arpoton

am unzuverlässigsten :

5. Quiromil

0 10 20 30 40 50 60 70 80 90 100 00 %

0 10 20 30 40 50 60 70 80 90 100 00 %

0 10 20 30 40 50 60 70 80 90 100 00 %

0 10 20 30 40 50 60 70 80 90 100 00 %

0 10 20 30 40 50 60 70 80 90 100 00 %

Weiter < Nächste Frage




Abbildung 7.20c: Erhebung der Testpräferenzen, Aufgabe 1b: Mit Hilfe der Schieberegler soll für jeden einzelnen Test dessen Validität eingeschätzt werden.

Fragen 1 2 3

1. Relative Häufigkeit korrekter Testergebnisse ?

Angenommen, Du würdest jeden einzelnen Test so oft hintereinander anwenden, bis dass er insgesamt 100-mal angesprochen hat, welcher der 5 Tests würde dann am häufigsten, zweit-häufigsten, etc. ein zuverlässiges Ergebnis liefern ?

am zuverlässigsten :

1. Waramux

2. Seperflen

3. Flogalam

4. Arpoton

am unzuverlässigsten :

5. Quiromil

0 10 20 30 40 50 60 70 80 90 100 84 %

0 10 20 30 40 50 60 70 80 90 100 72 %

0 10 20 30 40 50 60 70 80 90 100 65 %

0 10 20 30 40 50 60 70 80 90 100 43 %

0 10 20 30 40 50 60 70 80 90 100 33 %

Weiter < Nächste Frage




Abbildung 7.20d: Erhebung der Testpräferenzen, Aufgabe 1b: Einschätzung der Validitäten der Tests mit Hilfe der Schieberegler (Beispiel).

Fragen 1 2 3

2. Absolute Häufigkeit korrekter Testergebnisse ?

Angenommen, Du würdest jeden einzelnen Test genau 100-mal hintereinander anwenden, welcher der 5 Tests würde dann am häufigsten, zweit-häufigsten, etc. ein zuverlässiges Ergebnis liefern ?

am zuverlässigsten :

1.
2.
3.
4.
5.

am unzuverlässigsten :

1.
2.
3.
4.
5.




Abbildung 7.20e: Erhebung der Testpräferenzen, Aufgabe 2a: Die Tests sollen in eine Rangreihe gebracht werden. Das zu berücksichtigende Kriterium ist die Usefulness der Tests.

Fragen 1 2 3

2. Absolute Häufigkeit korrekter Testergebnisse ?


Angenommen, Du würdest jeden einzelnen Test genau 100-mal hintereinander anwenden, welcher der 5 Tests würde dann am häufigsten, zweit-häufigsten, etc. ein zuverlässiges Ergebnis liefern ?

am zuverlässigsten :

1.
2.
3.
4.
5.

am unzuverlässigsten :

1.
2.
3.
4.
5.



Angenommen, Du würdest den jeweils nebenstehenden Test (links) genau 100-mal hintereinander anwenden, wie oft (d.h. in wievielen dieser 100 Testanwendungen) würde er Dir dann ein zuverlässiges Ergebnis liefern ?

<input type="text" value="0"/>	<input type="text" value="10"/>	<input type="text" value="20"/>	<input type="text" value="30"/>	<input type="text" value="40"/>	<input type="text" value="50"/>	<input type="text" value="60"/>	<input type="text" value="70"/>	<input type="text" value="80"/>	<input type="text" value="90"/>	<input type="text" value="100"/>	<input type="text" value="00 %"/>
<input type="text" value="0"/>	<input type="text" value="10"/>	<input type="text" value="20"/>	<input type="text" value="30"/>	<input type="text" value="40"/>	<input type="text" value="50"/>	<input type="text" value="60"/>	<input type="text" value="70"/>	<input type="text" value="80"/>	<input type="text" value="90"/>	<input type="text" value="100"/>	<input type="text" value="00 %"/>
<input type="text" value="0"/>	<input type="text" value="10"/>	<input type="text" value="20"/>	<input type="text" value="30"/>	<input type="text" value="40"/>	<input type="text" value="50"/>	<input type="text" value="60"/>	<input type="text" value="70"/>	<input type="text" value="80"/>	<input type="text" value="90"/>	<input type="text" value="100"/>	<input type="text" value="00 %"/>
<input type="text" value="0"/>	<input type="text" value="10"/>	<input type="text" value="20"/>	<input type="text" value="30"/>	<input type="text" value="40"/>	<input type="text" value="50"/>	<input type="text" value="60"/>	<input type="text" value="70"/>	<input type="text" value="80"/>	<input type="text" value="90"/>	<input type="text" value="100"/>	<input type="text" value="00 %"/>
<input type="text" value="0"/>	<input type="text" value="10"/>	<input type="text" value="20"/>	<input type="text" value="30"/>	<input type="text" value="40"/>	<input type="text" value="50"/>	<input type="text" value="60"/>	<input type="text" value="70"/>	<input type="text" value="80"/>	<input type="text" value="90"/>	<input type="text" value="100"/>	<input type="text" value="00 %"/>

Abbildung 7.20f: Erhebung der Testpräferenzen, Aufgabe 2b: Mit Hilfe der Schieberegler soll für jeden einzelnen Test dessen Usefulness eingeschätzt werden.

Fragen
1 2 3

2. Absolute Häufigkeit korrekter Testergebnisse ?

Angenommen, Du würdest jeden einzelnen Test genau 100-mal hintereinander anwenden, welcher der 5 Tests würde dann am häufigsten, zweit-häufigsten, etc. ein zuverlässiges Ergebnis liefern ?

am zuverlässigsten :

1. Flogalam

2. Arpoton

3. Seperflen

4. Quiromil

am unzuverlässigsten :

5. Waramux

Angenommen, Du würdest den jeweils nebenstehenden Test (links) genau 100-mal hintereinander anwenden, wie oft (d.h. in wievielen dieser 100 Testanwendungen) würde er Dir dann ein zuverlässiges Ergebnis liefern ?

43 %

36 %

14 %

08 %

05 %

Weiter

<

Nächste Frage

Abbildung 7.20g: Erhebung der Testpräferenzen, Aufgabe 2b: Einschätzung der Usefulness der Tests mit Hilfe der Schieberegler (Beispiel).

Fragen
1 2 3

3. Ansprech-Häufigkeit der Tests ?

Angenommen, Du würdest jeden einzelnen Test genau 100-mal hintereinander anwenden, welcher der 5 Tests würde dann am häufigsten, zweit-häufigsten, etc. ansprechen ?

spricht am häufigsten an :

1. Quiromil

2. Arpoton

3. Seperflen

4. Flogalam

spricht am seltensten an :

5. Waramux

Weiter

Abbildung 7.20h: Erhebung der Testpräferenzen, Aufgabe 3a: Bildung einer Rangreihe gemäss dem Kriterium der Diskriminationsrate der Tests (Beispiel).

Fragen 1 2 3

3. Ansprech-Häufigkeit der Tests ?

Angenommen, Du würdest jeden einzelnen Test genau 100-mal hintereinander anwenden, welcher der 5 Tests würde dann am häufigsten, zweit-häufigsten, etc. ansprechen ?

spricht am häufigsten an : 1.

2.

3.

4.

spricht am seltensten an : 5.

Angenommen, Du würdest den jeweils nebenstehenden Test (links) genau 100-mal hintereinander anwenden, wie oft (d.h. in wievielen dieser 100 Testanwendungen) würde er dann überhaupt ansprechen (d.h. eine korrekte oder falsche Vorhersage machen) ?

Test	0	10	20	30	40	50	60	70	80	90	100	Percentage
Quiromil												85 %
Arpoton												50 %
Seperflen												42 %
Flogalam												37 %
Waramux												05 %

Q S F W A

Weiter < Nächste Frage

Abbildung 7.20i: Erhebung der Testpräferenzen, Aufgabe 3b: Einschätzung der Diskriminationsrate der Tests mit Hilfe der Schieberegler (Beispiel).

Hinweis zum Beginn einer neuen Expedition: Auf einer separaten Bildschirmseite künden wir jeweils den Beginn der nächsten Expedition an und fordern die Versuchspersonen auf, die neue Untersuchungsbedingung zu beachten (siehe Abbildungen 7.10, 7.13 und 7.16).

Anwendungsphase: Die Anwendungsphase besteht aus drei Expeditionen, in denen die Versuchspersonen nun möglichst viele Äffchen einer Wildpopulation retten sollen. Sie können dazu ihre Erfahrung mit den fünf Tests einsetzen, um bei jeweils 12 eingefangenen Äffchen zu entscheiden, welche davon mit dem Medikament zu behandeln sind und welche nicht. Anders als in der Lernphase wählen sie dazu nun selbst sequentiell Tests aus. Pro Test steht eine eigene Schaltfläche zur Verfügung. Wird diese angeklickt, erscheint das Testergebnis auf dem Bildschirm. Allerdings sind die Versuchspersonen hinsichtlich des Testeinsatzes in jeder Expedition an eine spezifische Untersuchungsbedingung gebunden, die den Abbruch der sequentiellen Testserie steuert. Die Instruktionen für die drei Expeditionen sind im Anhang (siehe Anhänge, 7.8.2) im Wortlaut wiedergegeben. Bei korrekter Handhabung führen die Untersuchungsbedingungen zur Anwendung der Validitätsordnung, des Success und der Diskriminationsrate. Spätestens dann, wenn eine Versuchsperson keine weiteren Tests mehr durchführen kann, muss sie sich entscheiden, ob sie dem Äffchen das Medikament applizieren will oder nicht. Die Abbildungen 7.11a–7.11e (Expedition 1), 7.14a–7.14c (Expedition 2) sowie 7.17a–7.17d (Expedition 3) zeigen exemplarisch den Ablauf eines Durchgangs für jede Expedition.

Ende einer Expedition: Ist das Ende einer Expedition erreicht, weisen wir mit einer entsprechenden Hinweisseite darauf hin (Abbildungen 7.12, 7.15, 7.18).

Befragung: Anschliessend an die dritte (und letzte) Expedition folgt der aus drei Aufgabenpaaren bestehende Befragungsteil zur Bewertung der Cue-Güte, den wir wiederum mit einer Hinweisseite ankündigen (siehe Abbildung 7.19). In allen drei Aufgabenpaaren müssen die Versuchspersonen jeweils anhand eines vorgegebenen Gütekriteriums eine Rangreihe der fünf Tests bilden und anschliessend für jeden Test mittels Schieberegler dessen Güte prozentgenau einschätzen. In Aufgabe 1a müssen die Versuchspersonen die Rangreihe gemäss der wahrgenommenen Zuverlässigkeit (Validität) bilden und in Aufgabe 1b dann für jeden Test die absolute Ausprägung der Zuverlässigkeit einschätzen (Abbildungen 7.20a–7.20d). Dasselbe geschieht anschliessend auch für die Usefulness³⁶ in Aufgabenpaar 2a/2b (Abbildungen 7.20e–7.20g) und die Diskriminationsrate in Aufgabenpaar 3a/3b (Abbildungen 7.20h–7.20i). So korrespondiert also jedes Aufgabenpaar mit einer der drei Expeditionen der Anwendungsphase, wobei die Aufgabenstellungen jeweils ohne direkten Bezug zur entsprechenden Expedition formuliert sind (siehe Anhänge, 7.8.3).

Anschliessend folgt die Nachbefragung, auf die wir mit einer gesonderten Bildschirmseite hinweisen (Abbildung 7.21).

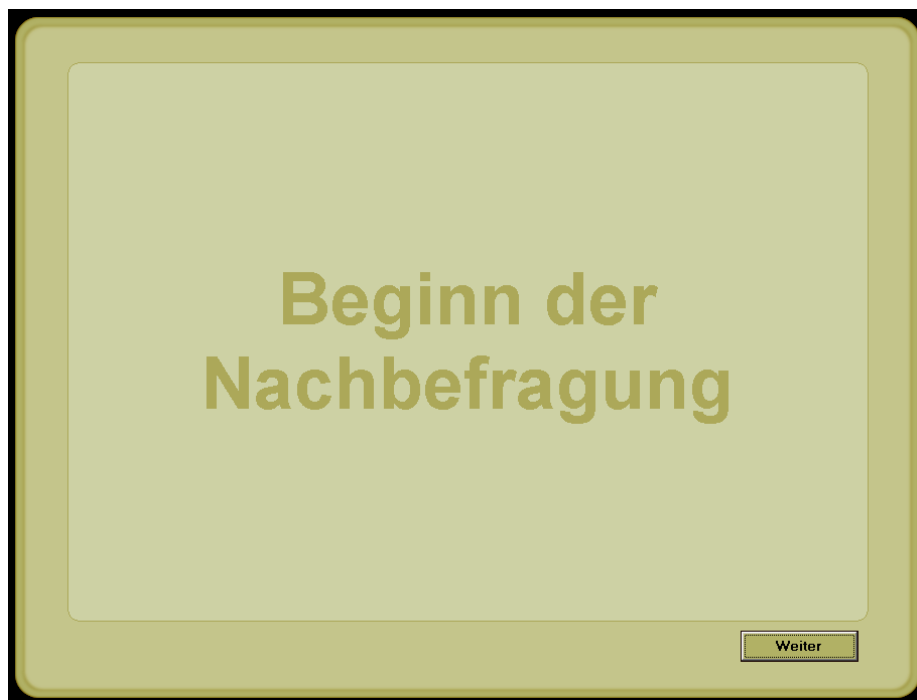


Abbildung 7.21: Hinweis auf den Beginn der Nachbefragung

³⁶ Wie bereits in der Einleitung erwähnt, sind Verhaltensdaten und Befragungsdaten hinsichtlich Success (Verhalten) und Usefulness (Befragung) dissoziiert. Dies erlaubt Aussagen über beide integrierte Gütemasse und somit eine Erweiterung der referierten Erkenntnisse von Newell et al. (2003). Mit der Befragung sollen also nicht in erster Linie die Resultate der Anwendungsphase überprüft werden, sondern das Ziel besteht in einer breiteren Datenbasis für das bei den Vpn vorhandene Verständnis der Cuegüte. Aufgrund der Instruktion für Expedition 2 (die Success als Norm implizierte) umfasste die aus dieser Expedition resultierende Datenspur jeweils lediglich einen Cue pro Durchgang. Wichtig war uns deshalb in der Befragung, dass die Vpn nicht nur einen einzigen Cue als integriertes Gütemass auswählen, sondern sämtliche 5 Cues bezüglich einer Suchstrategie rangieren, die V und D kombiniert.

Nachbefragung: Für die Nachbefragung präsentieren wir – getrennt nach Expedition – sowohl die von der Versuchsperson in jeder Runde konsultierten Tests als auch die im Befragungsteil erhobenen Rangreihen gemeinsam auf einer Bildschirmseite. Auf derselben Bildschirmseite geben wir zum Vergleich auch die normativ korrekten Entscheidungen und Rangreihen bekannt (siehe Abbildung 7.22). Anhand dieser Informationen wird der Versuchsleiter kurz das Informationssuchverhalten der Versuchsperson analysieren und sie anschliessend darüber aufklären, was wir untersuchen, welche Bedeutung die dargestellten Daten in diesem Zusammenhang haben, wie die jeweils beste Lösung aussieht und wie diese Zustände kommt. Der Versuchsleiter klärt die Versuchsperson also zunächst über das von ihr in den drei Expeditionen gezeigte Informationssuchverhalten auf. Anschliessend wird er anhand eines halbstrukturierten Interviews versuchen herauszufinden, welche Überlegungen die Versuchsperson bei der Auswahl der Tests ange stellt hat und erfasst diese ggf. über das zur Verfügung stehende Textfeld. Zuunterst auf der Seite geben wir schliesslich noch das in den drei Exkursionen erzielte Gesamtergebnis, d.h. den Anteil insgesamt überlebender Äffchen, bekannt.

Nachbefragung

Exkursion 1

Testauswahl für Zuverlässigkeit [Z]

	Warami	Sepette	Flogalar	Apotoni	Quiromi
1	1	2	3		
2	1	2			
3	1	2	3	4	
4	1	2	3	4	
5	1	2	3		
6	1	2	3		
7	1				
8	1	2	3	4	5
9	1	2			
10	1	2	3		
11	1	2	3		
12	1	2	3	4	
Rang	1	2	3	4	5
VP	Wert	Wert	Wert	Wert	Wert
VP	84	72	65	43	33
eff.	100	82	82	64	50

Anmerkungen zur Testauswahl

Exkursion 2

Testauswahl für Kombination [A*Z]

	Flogalar	Quiromi	Apotoni	Sepette	Warami
1	1				
2	1				
3	1				
4	1				
5	1				
6	1				
7	1				
8	1				
9	1				
10	1				
11	1				
12	1				
Rang	1				
VP	Wert	Wert	Wert	Wert	Wert
VP	43	8	36	14	5
eff.	50	40	39	18	10

Anmerkungen zur Testauswahl

Exkursion 3

Testauswahl für Ansprechrate [A]

	Quiromi	Apotoni	Flogalar	Sepette	Warami
1	4	1	3	5	2
2		1			
3		1			
4		1			
5		1			
6		1			
7		1			
8		1			
9		1			
10		1			
11		1			
12		1			
Rang	1	2	3	4	5
VP	Wert	Wert	Wert	Wert	Wert
VP	85	50	37	42	5
eff.	80	61	61	22	10

Anmerkungen zur Testauswahl

>>> Anteil überlebender Äffchen : 83.3 % (12 + 10 + 8 = 30 von Total 36) <<<

Weiter

Abbildung 7.22: Nachbefragung.

Persönliche Angaben: Die Ankündigung erfolgt wiederum mit einer gesonderten Bildschirmseite (Abbildung 7.23). Mit der folgenden Seite (Abbildung 7.24) erfassen wir dann die persönlichen Angaben zur Versuchsperson (Kontaktinformationen, Beruf, Geburtsdatum, Geschlecht), klären die Bereitschaft zur Teilnahme an allfälligen Nachfolge-Experimenten ab und protokollieren auf Wunsch allfällige Kommentare zum Experiment.

Bedankung: Auf dieser Seite bedanken wir uns für die Teilnahme am Experiment. Dies ist die letzte Seite des Experiments (siehe Abbildung 7.25).

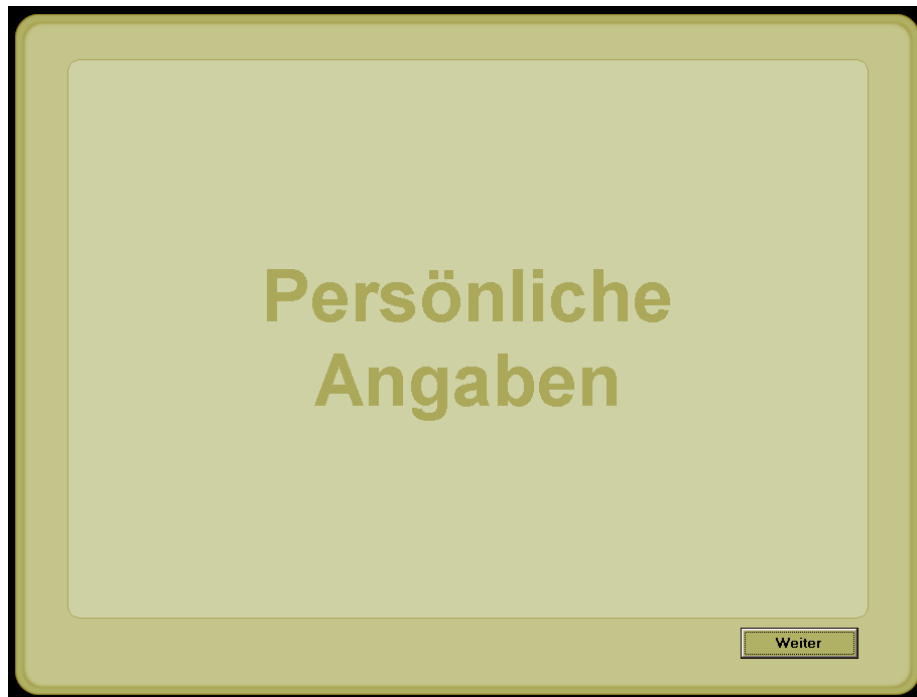


Abbildung 7.23: Hinweis auf die folgende Erfassung persönlicher Angaben.

Persönliche Angaben

Adresse & demographische Daten

Vorname: *	<input type="text"/>	Beruf (bzw. Studienfach / Semester): *	<input type="text"/>
Nachname: *	<input type="text"/>	Geburts-Datum: *	<input type="text"/> <input type="text"/> <input type="text"/> Geschlecht: * <input type="text"/>
Strasse + Nr. *	<input type="text"/>	Telefon-Nummer:	<input type="text"/>
PLZ + Ort: *	<input type="text"/>	Email-Adresse:	<input type="text"/>

Mit einem Stern (*) markierte Felder erfordern zwingend eine Eingabe

Teilnahme an zukünftigen Experimenten

☒ Ich bin bereit, mich für ein (allfälliges) Nachfolge-Experiment zur Verfügung zu stellen.
☐ Ich möchte in Zukunft nicht mehr an Experimenten an eurer Abteilung teilnehmen.

Grund (Angabe optional):

Kommentare

Mir ist bei eurem Experiment folgendes aufgefallen: / Vorschläge, Anmerkungen, etc:

Abbildung 7.24: Maske zur Erfassung persönlicher Angaben.

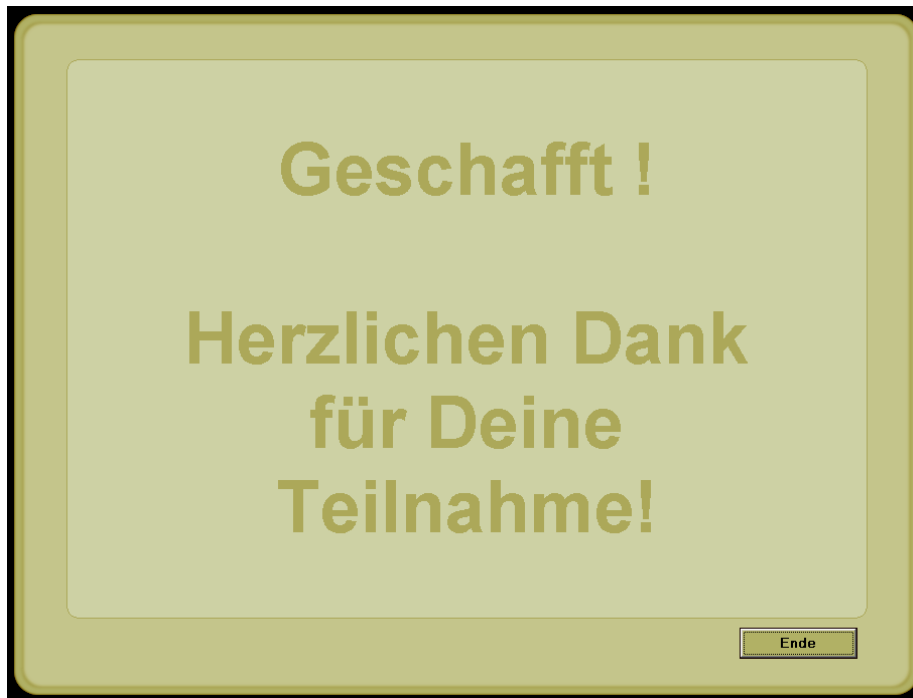


Abbildung 7.25: Auf der letzten Seite bedanken wir uns für die Teilnahme am Experiment.

7.4 Datengrundlage

Mit der Konstruktion des Experiments haben wir für jede Aufgabe festgelegt, welche³⁷ Informationen von den fünf Cues angezeigt werden (siehe Anhänge, 7.8.4). Diese Festlegungen gelten für alle Versuchspersonen gleichermassen. Um möglichen Salienzeffekten eines Testnamens entgegenzuwirken, variieren wir die Zuordnung der Testnamen zu den Cues von Versuchsperson zu Versuchsperson. Innerhalb eines Blocks von 10 Aufgaben (Lernphase) bzw. 12 Aufgaben (Testphase) sind die Validitäten und Diskriminationsraten repräsentativ verteilt. Innerhalb dieser Blöcke wird die Reihenfolge der Aufgaben (zur Laufzeit des Experiments) zufällig bestimmt.

Cue-Struktur: Tabelle 7.3 zeigt die Zuordnung der Gütemasse zu den Cues. Die Cue-Struktur ist folgendermassen aufgebaut: Cue 1 führt immer zum korrekten Ergebnis, diskriminiert aber nur in 10% der Fälle. Die Cues 2 und 3 haben eine gute Validität (.82), unterscheiden sich aber markant in ihren

Tabelle 7.3: Die fünf Cues und ihre Gütwerte

Cue-Nr	V	D	Success	Usefulness
1	1.0	.10	.55	.10
2	.82	.22	.57	.18
3	.82	.61	.70	.50
4	.64	.61	.59	.39
5	.50	.80	.50	.40

³⁷ Genaugenommen haben wir nicht definiert, ob ein „+“ oder ein „-“, angezeigt werden soll, sondern ob die angezeigte Information korrekt („2“) oder falsch („1“) ist. eine Null in der Tabelle bedeutet, dass der Cue in der entsprechenden Aufgabe nicht diskriminiert (und dementsprechend ein „?“ angezeigt wird).

Diskriminationsraten (.22 vs. .61). Die Cues 3 und 4 haben beide eine hohe Diskriminationsrate (.61), jedoch sehr unterschiedliche Validitäten (.82 vs. .64). Cue 5 schliesslich zeichnet sich durch eine extrem hohe Diskriminationsrate (.80) aus, ist im Hinblick auf die Validität jedoch völlig nutzlos (.50). Die Validitäten und Diskriminationsraten verlaufen absichtlich gegenläufig und sind so gewählt, dass sich für Cue 3 der beste Success (.70) ergibt.

7.5 Benutzeroberflächen

In diesem Kapitel gehen wir auf die wichtigsten Überlegungen ein, die bei der Entwicklung der verschiedenen Benutzeroberflächen des Experiments eine wichtige Rolle gespielt haben.

7.5.1 Login

Die Benutzeroberfläche des Anmeldebildschirms enthält im Wesentlichen lediglich zwei drop-down Listenfelder zur Erfassung des Namens des Versuchsleiters bzw. des Namens (oder der anonymisierten Bezeichnung) der Versuchsperson. Beide Felder lassen sowohl die Auswahl eines bereits erfassten Namens als auch die Neueingabe eines Namens oder einer Bezeichnung zu. Im drop-down Bereich des Listenfelds „Versuchsperson“ taucht der Name einer bereits erfassten Versuchsperson jedoch nur dann auf, wenn diese das Experiment noch nicht vollständig beendet hat. Sollte dies jemals der Fall sein, so wird das Experiment durch die Auswahl des entsprechenden Namens an derselben Stelle fortgesetzt, an der es zuvor unterbrochen wurde. Unterkapitel 7.6 („Details zur Programmierung“, Abschnitt „Unterbrechbarkeit“) enthält weitere Ausführungen zur Möglichkeit, mit dem Experiment nach einem Unterbruch an derselben Stelle fortzufahren. Um zu verhindern, dass das Experiment ohne Angabe eines Versuchspersonennamens gestartet werden kann, wird die Schaltfläche „Weiter“ erst nach erfolgter Eingabe (bzw. Auswahl) eines Versuchspersonennamens aktiviert. Zum Login sei an dieser Stelle noch vermerkt, dass eine Versuchsperson in der Regel bereits den Anmeldebildschirm zu Gesicht bekommt. Deshalb darf dieser keine Informationen enthalten, die einen Rückschluss auf die zu untersuchende Fragestellung zulassen, was bspw. bei einem ungünstig gewählten Experiment-Titel denkbar wäre. Dasselbe gilt natürlich auch für alle nachfolgenden Bildschirmseiten. Der im Anmeldebildschirm aufgeführte Titel des Experiments lautet deswegen schlicht und neutral „Experiment“.

7.5.2 Lern- und Anwendungsphase

Die folgenden Abschnitte gelten, sofern nicht anderweitig vermerkt, sowohl für die Lernphase als auch für die Anwendungsphase des Experiments.

Kompaktes „All-In-One-Design“: Alle für die Versuchspersonen relevanten Informationen werden direkt auf dem Bildschirm präsentiert, so dass keine mündlichen Erklärungen zum Versuchsablauf abgegeben werden müssen und sämtliche Versuchsmaterialien (insbesondere der Instruktionen) in standardisierter Form vorliegen bzw. auch zu einem viel späteren Zeitpunkt (d.h. nach Monaten oder Jahren) noch in identischer Weise vorliegen werden. Dank der so erreichten Kompaktheit des Experiments bzw. des Versuchsmaterials kann dieses zudem auf unkomplizierte Weise weitergegeben und archiviert werden.

Titel und Schriftgrößen: Die einzelnen Bildschirmseiten sind zwecks klarer Orientierung über den Fortgang des Experiments mit Seitentiteln („Lernphase“, „Exkursion“ etc.) versehen und sämtliche auf der Benutzeroberfläche vorhandenen Textelemente sind in gut lesbaren Schriftgrößen dargestellt.

Panels zur Gruppierung von Elementen: Zusammengehörige Elemente (wie bspw. Schaltflächen, Labels, Icons, etc.) der einzelnen Experimental-Schritte sind in insgesamt 4 Panels gruppiert. Die Gruppierung führt zu einer klareren optischen Abgrenzung zwischen den einzelnen Experimental-Schritten.

Hintergrund und Farbgebung: Der unaufgeregte, einfarbige Hintergrund sowie die bündige Ausrichtung der auf der Oberfläche platzierten Elemente an einem unsichtbaren Raster, sollen zur Minimierung unerwünschter Distaktionen von der eigentlichen Aufgabe beitragen. Der Oberflächen-Hintergrund sowie die Rahmen und Innenflächen der einzelnen Panels sind in abgestuften, ruhigen Blautönen gehalten. Dabei nimmt die Helligkeit von aussen nach innen (d.h. vom Rahmen der Benutzeroberfläche bis hin zum Hintergrund eines Panels) stetig zu, was zu einer optischen Fokussierung auf die eigentlichen Aktions- bzw. Interaktionselemente führt.

Abarbeitungsrichtung der Experimental-Schritte: Die Bildschirmoberfläche ist so gestaltet, dass das Abarbeiten der Versuchsdurchgänge jeweils „von links nach rechts“ bzw. „von oben nach unten“ erfolgt. Dank dieser bereits aus dem Alltag gewohnten Leserichtungen sollen sich die Versuchspersonen innerhalb der Durchgänge möglichst einfach orientieren können. Die Abarbeitungsreihenfolge der Experimental-Schritte ist durch die den Panel-Titeln voran gestellten Schritt-Nummern eindeutig festgelegt. Die Nummerierung der einzelnen Experimental-Schritte ermöglicht uns, eindeutige Bezüge zwischen den in der oberen Bildschirmhälfte platzierten Instruktionen und den darunterliegenden Panels herzustellen.

Grösse und Anordnung der Schaltflächen: Bei den Schaltflächen wurde darauf geachtet, diese erstens nicht zu klein zu dimensionieren, damit man sie auch ohne umständliche Mauszeiger-Positionierungsübungen leicht trifft und zweitens genügend Abstand zwischen nebeneinander liegenden Schaltflächen vorhanden ist, damit bei hastigem bzw. ungenauem Positionieren des Mauszeigers ein ungewolltes Anklicken eines benachbarten Steuerelements möglichst verhindert wird.

Keine keyboard shortcuts: Das ganze Experiment basiert auf einer Maussteuerung und kann bzw. muss – von einigen wenigen Eingaben bei der Nachbefragung abgesehen – ausschliesslich mit der Maus bedient werden. Interaktionen mit Steuerelementen mittels Tastenkombinationen (sog. keyboard shortcuts) sind nicht möglich. Damit lassen wir erst gar keine Möglichkeit zu, dass sich die Versuchspersonen bei ihren Eingaben vertippen können und so bspw. ungewollt Cues auswählen die sie eigentlich gar nicht hätten konsultieren wollen.

Information-Board: Für die Lern- und die Testphase haben wir ein Information-Board Design (in Anlehnung an das Mouselab von Payne, Bettman & Johnson, 1993) verwendet. Im Gegensatz zur üblicherweise verwendeten tabellarischen Darstellungsform haben wir die Cues in diesem

Experiment allerdings äquidistant auf einer gedachten Ellipse auf dem Bildschirm angeordnet, um den Eindruck einer festen Reihenfolge möglichst zu minimieren.

Repräsentation der Cues: Bei den Cues handelt es sich um medizinische Tests, deren Namen wir frei erfunden haben (um Vorwissen auszuschliessen), die unterschiedliche Anfangsbuchstaben haben (zwecks besserer Differenzierung beim Erlernen der Cuegüte) und deren Silbenzahl identisch ist (um keine Zusammenhangs-Hypothesen zwischen der Silbenlänge und der Cuegüte zu evozieren und einzelnen Cues aufgrund ihrer Silbenlänge nicht ungewollt zu einer höheren Salienz zu verhelfen).

Farbgebung der Cue-Symbole: Beim Cue-Lernen sollen die Versuchspersonen rasch einen möglichst adäquaten Eindruck der Auftretenshäufigkeiten bestimmter Ereignisse gewinnen. Das relevante Ereignis beim Erlernen der Diskriminationsraten lautet: „Der Test hat ein Resultat geliefert!“ (d.h. „+“ oder „–“, jedoch nicht „?“). Beim Erlernen der Validitäten lautet das relevante Ereignis: „Der Test hat ein korrektes Resultat geliefert!“. Für das Cue-Lernen ist in beiden Fällen entscheidend, dass die Ereignisse möglichst gut mit dem jeweils durchgeführten Test (und nur mit diesem) verknüpft werden. Durch die Zuordnung unterschiedlicher Farben haben wir deshalb die optische Unterscheidbarkeit (Salienz) der Tests erhöht. Die Farben sind augenscheinlich so ausgewählt, dass sie, paarweise miteinander verglichen, möglichst gut voneinander unterschieden werden können. Nachdem ein Cue konsultiert worden ist, wird er farblich anders dargestellt als vor seiner Konsultation: Ein Cue der nicht diskriminiert, wechselt von seiner ursprünglichen Farbe auf die Hintergrundfarbe des Panels (ein helles grau-blau), tritt damit also wörtlich in den Hintergrund. Ein Cue der diskriminiert, behält zwar seine Farbe, wird allerdings mit einer erhöhten Farbsättigung dargestellt und tritt damit (gegenüber den noch nicht konsultierten Cues) etwas stärker in den Vordergrund. Wir gehen davon aus, dass sich die grössere Salienz der Cues positiv – im Sinne einer höheren Lerngeschwindigkeit – auf den Prozess des Cue-Lernens auswirkt.

Aktive Suche nach Cue-Information (Konsultation von Cues): In den drei Anwendungsphasen sind die Cues (bzw. die Bezeichnungen der Tests) als Schaltflächen implementiert. Dementsprechend müssen die Versuchspersonen bei der Informationssuche mit dem Mauszeiger direkt auf die einzelnen Testnamen klicken, um an die Informationen der entsprechenden Cues zu gelangen. Auch in der Lernphase müssen die Cue-Informationen aktiv gesucht werden. Allerdings muss hier lediglich eine einzige Schaltfläche („Testresultate anzeigen“, rechts unten im Information-Board Panel) angeklickt werden.

Darstellung der Cue-Informationen: Bei den Symbolen die wir zur Darstellung der Cue-Informationen verwenden, orientieren wir uns an der im medizinischen Bereich üblichen Symbolik, d.h.: „+“ für Test hat angeschlagen und „–“ für Test hat nicht angeschlagen. Mittels eines Fragezeichens („?“) zeigen wir an, dass ein Test nicht funktioniert hat und dementsprechend weder eine positive noch eine negative Aussage zulässt.

Räumliche Nähe zwischen Cues, Cue-Informationen und Feedback: Während für das Erlernen der Diskriminationsraten bereits die Anzeige der Cue-Informationen (Testresultate) in symbolischer Form („+“, „–“, „?“) ausreicht, werden für das Erlernen der Validitäten zusätzlich die

Rückmeldungen über die Korrektheit der angezeigten Testresultate benötigt. Die Rückmeldungen werden direkt an Ort-und-Stelle der entsprechenden Objekte (d.h. bei der visuellen Repräsentation der einzelnen Tests) eingeblendet. Dabei wird ein korrektes Testresultat mit einem Häckchen versehen und ein falsches mittels eines Kreuzes durchgestrichen. Wir gehen davon aus, dass sich die räumliche Nähe zwischen den Testnamen, Testresultaten und Rückmeldungen vorteilhaft auf die Effizienz beim Cue-Lernen auswirkt.

Darstellung eines Referenzobjektes: Im Zentrum der elliptisch angeordneten Cues (Tests) wird in allen Runden dasselbe verfremdete Bild eines Referenzobjektes (Äffchen) aus der Referenzmenge (Äffchenpopulation) angezeigt. In der Instruktion erwähnen wir, dass es sich dabei um sogenannte „Tigerschwanzäffchen“ handelt. Wir haben diese Bezeichnung frei erfunden, um den Abruf von Vorwissen über eine bestimmte Affenart möglichst zu minimieren. Aus dem gleichen Grund wurde deshalb das Bild des dargestellten Äffchens (einer real noch existierenden Affenart) visuell bearbeitet (Farbinversion, Unschärfefilter, Freistellung).

Effektiver Zustand der Zielvariablen: Der Ausprägungsgrad der Zielvariablen (d.h. der effektive Gesundheitszustand des Äffchens) wird durch ein entsprechendes Symbol („+“ für krank, „-“ für gesund) direkt bei der Darstellung des Referenzobjektes, dem Äffchen-Bild, eingeblendet. Für das Erlernen der Validitäten ist diese Rückmeldung jedoch nicht zwingend erforderlich. Das Verhältnis von kranken zu gesunden Äffchen beträgt 1:1 und wird den Versuchspersonen bereits in der Instruktion bekanntgegeben, um sie von der Verwendung einer Ratestrategie abzuhalten.

Visuelle Repräsentation der Konsequenz einer Entscheidung: Anhand von Emoticons (Smileys) sollen die Versuchspersonen auf einen Blick die Folge einer Entscheidung auf dem Bildschirm ablesen können. Zusätzlich zum Emoticon wird jeweils ein kurzer Text eingeblendet, der die Konsequenz der Entscheidung in schriftlicher Form erläutert. Der Text gibt zudem Auskunft über den Ausprägungsgrad der Zielvariablen (d.h. den tatsächlichen Gesundheitszustand des Äffchens, den dieses vor der Entscheidung innehatte).

Die Bildschirmoberflächen der drei Teile der Anwendungsphase sind prinzipiell identisch mit derjenigen der Lernphase. Die wenigen zwischen der Lern- und der Anwendungsphase bestehenden Unterschiede, werden in den nachfolgenden Abschnitten genauer erläutert:

Suche nach Cue-Information: Ein erster Unterschied zwischen der Lern- und der Anwendungsphase besteht, wie bereits weiter oben erwähnt, darin, dass die Versuchspersonen die fünf Cues in der Lernphase nicht separat konsultieren können, um an deren Informationen zu gelangen. Durch das Anklicken einer einzigen, rechts unten im Panel platzierten Schaltfläche, erhalten sie bereits die Hinweise sämtlicher Cues. Damit fokussieren wir die Versuchspersonen schon in der Lernphase auf die aktive Informationssuche und sorgen gleichzeitig für optimale Lernbedingungen, weil die Versuchspersonen in jedem Durchgang sofort sehen, welche Cues diskriminieren und welche nicht bzw. welche (der diskriminierenden) Cues einen korrekten Hinweis geliefert haben. In der Anwendungsphase hingegen können sie die Cues nur in serieller Reihenfolge konsultieren, weil wir dort herausfinden wollen, welche Cues sie überhaupt auswählen und in welcher Reihenfolge sie dies tun. Im Zusammenspiel mit geeigneten Instruktionsvarianten können

wir dadurch ausserdem verschiedene Informationssuch-Strategien testen und dabei ermitteln, wie gut die Validitäten und Diskriminationsraten der Cues gelernt wurden.

Konditionale Formulierungen: Der zweite Unterschied zwischen den beiden Teilen besteht darin, dass alle Formulierungen in Panel 2 („Wie würdest Du entscheiden?“ bzw. „Wie entscheidest Du Dich?“) sowie die Erläuterungstexte in Panel 3 („Folge Deiner Entscheidung“) in der Lernphase konditional formuliert sind, weil die Versuchspersonen den Äffchen das Medikament während des Trainings nicht wirklich verabreichen sondern lediglich einen hypothetischen Behandlungsentscheid fällen müssen (siehe Instruktion).

7.5.3 Befragung

Im Befragungsteil verwenden wir für alle drei Aufgabenpaare grundsätzlich dieselbe Oberfläche. Abgesehen von den Inhalten der beiden Fragefelder ändert zwischen den einzelnen Aufgaben lediglich die zur Rangreihenbildung angegebene Skala. Die Oberfläche ist horizontal in zwei gleich grosse Bereiche unterteilt: Die linke Hälfte enthält fünf vertikal entlang der Skala angeordnete drop-down Listenfelder zur Rangreihenbildung, wobei lediglich die beiden Pole der Skala angegeben sind. Die Rangreihe muss aus den fünf bereits bekannten Cues zusammengesetzt werden. Das dabei zu verwendende Gütekriterium ist im Fragefeld (oberhalb des ersten Listenfelds) genau beschrieben. Links unterhalb des letzten Listenfelds ist als Erinnerungshilfe zudem ein verkleinertes Abbild des in der Expeditionsphase präsentierten Information-Boards zu finden. Die rechte Hälfte enthält fünf Schieberegler zur prozentgenauen Angabe der geschätzten Gütewerte der Cues. Die Schieberegler sind vertikal und korrespondierend zu den auf der linken Hälfte liegenden Listenfeldern angeordnet. Das bei der Einschätzung zu berücksichtigende Kriterium ist auch hier in einem Fragefeld (oberhalb des ersten Schiebereglers) detailliert beschrieben.

In den folgenden Abschnitten gehen wir näher auf einige technische Details des Befragungsteils ein:

Abarbeitungsrichtung und Zweiteilung der Oberfläche: Auch in dieser Oberfläche werden die einzelnen Experimentalschritte wieder in gewohnter Manier, d.h. von links nach rechts bzw. von oben nach unten abgearbeitet. Um eine Versuchsperson Schritt für Schritt durch die Befragung zu führen und dabei Distractionen möglichst zu vermeiden, wird die rechte Seite der Oberfläche erst eingeblendet, wenn die Versuchsperson mit der Eingabe der Rangreihe fertig ist. Erst dann kann sie auf der rechten Seite mit der Einschätzung der Güte der einzelnen Cues beginnen.

Rückkehr zur Rangreihenbildung: Solange eine Versuchsperson die Einschätzung der Güte der Cues nicht abgeschlossen hat, kann sie jederzeit wieder auf die linke Seite zur Rangreihenbildung zurückkehren – bereits erfasste Einschätzungen gehen dabei allerdings verloren (und die rechte Seite wird ausgeblendet). Während zur Bildung einer Rangreihe bereits eine einigermaßen vage Vorstellung von der Güte der einzelnen Cues ausreicht, müssen bei der konkreten Einschätzung der Güte prozentgenaue Werte angegeben werden. Dies könnte die Entstehung einer anderen als der zuvor angegebenen Cue-Rangreihe zur Folge haben. Deshalb ist es wichtig, dass

eine Versuchsperson ggf. zur Rangreihenbildung zurückkehren kann, um die zuvor eingegebene Rangreihe abzuändern.

Dynamische drop-down Listenfelder: Die auf der linken Seite verwendeten Listenfelder haben zwei dynamische³⁸ Eigenschaften: Der drop-down-Bereich (die Liste) jedes Listenfelds ist zu Beginn der Befragung mit den Namen aller fünf Tests sowie einem leeren Eintrag bestückt. Die Reihenfolge der Tests ist zufallsgeneriert und bei allen fünf Listen identisch. Der leere Eintrag steht immer zuoberst in der Liste. Sobald eine Versuchsperson einen Test auswählt, wird dieser aus sämtlichen Listen entfernt. Dadurch ist auf einfache Weise sichergestellt, dass kein Test mehrfach ausgewählt werden kann. Ersetzt eine Versuchsperson in einem Listenfeld einen bereits ausgewählten Test durch einen anderen Test oder durch den leeren Listeneintrag, so wird der soeben ersetzte Test wieder allen fünf Listen hinzugefügt. 2) Mit der Auswahl eines Tests wechselt im entsprechenden Listenfeld die Farbe des Hintergrunds: Anstelle von weiss nimmt dieser die Farbe des Tests an – das ist diejenige Farbe, die dem betreffenden Test bereits im Information-Board zugeordnet war. Die farbliche Markierung der ausgewählten Tests bietet den Versuchspersonen eine zusätzliche Möglichkeit zur Überprüfung ihrer Auswahl. Sie sollen auf einen Blick feststellen können, ob sie denjenigen Test ausgewählt haben, den sie tatsächlich haben auswählen wollen.

Dimensionierung der Schieberegler: Die Schieberegler müssen eine geeignete Länge haben, damit bei der grafischen Darstellung keine störenden Interpolationseffekte auftreten. Eine geeignete Länge (gemessen in Anzahl Pixel von und mit Skalenstrich „0“ bis und mit Skalenstrich „100“) ist dann gefunden, wenn sämtliche nachfolgend genannten Bedingungen erfüllt sind: 1) Die Skalenstriche erscheinen auf dem Bildschirm pixelgenau äquidistant; 2) Der Reglerknopf (schwarzes Dreieck) legt bei jeder Verschiebung um eine Einheit ($\pm 1\%$) die gleiche Distanz, gemessen in Anzahl Pixel, zurück – unabhängig davon, an welcher Stelle er sich gerade befindet; 3) Für jeden einzelnen Prozentwert zwischen 0 und 100 gibt es genau eine eindeutige Reglerposition; 4) Die Spitze des Reglerknopfes ist genau ein Pixel breit und zeigt immer genau auf den eingestellten Wert (bzw. auf einen Skalenstrich, sofern vorhanden). Pixelbreiten (B) die alle genannten Bedingungen erfüllen, können gemäss der folgenden Formel berechnet werden: $B = 101 + n \cdot 100$ (wobei $n \geq 0$). Die 4 Bedingungen sind also genau dann erfüllt, wenn $B = \{101; 201; 301; \dots\}$. Eine Abschätzung anhand des Kriteriums „Verfügbarer Platz in der rechten Hälfte der Bildschirmoberfläche“ ergab eine optimale Breite von 301 Pixel. Mit einer Breite von 401 Pixel hätte der Regler bereits nicht mehr in die rechte Oberflächenhälfte gepasst und mit 201 Pixel wäre er kleiner gewesen als zwingend nötig und damit weniger gut bedienbar.

7.5.4 Nachbefragung

Die Nachbefragung steht in Bezug zu den drei Expeditionen der Anwendungsphase. Die Oberfläche ist dementsprechend (in horizontaler Richtung) dreigeteilt. Die einzelnen Drittel wiederum sind im Wesentlichen in zwei Bereiche unterteilt. Im DataGrid-Element im oberen Bereich

³⁸ dynamisch im Sinne einer Anpassung an veränderte Bedingungen.

zeigen wir jeweils alle in einer Expedition konsultierten Cues sowie die zur Expedition korrespondierende, im Befragungsteil erhobene, Cue-Rangreihe an und präsentieren zum Vergleich die normativ korrekten Entscheidungen und die normativ korrekte Cue-Rangreihe. Das Textfeld im unteren Bereich bietet ausreichend Platz zur Erfassung allfälliger Anmerkungen zum Informationsverhalten.

Bei der Entwicklung der Oberfläche waren in erster Linie die beiden Kriterien Kompaktheit und Übersichtlichkeit massgebend:

Kompaktheit: Alle für das Debriefing bzw. die Nachbefragung benötigten Informationen sollen auf einer einzigen Bildschirmseite untergebracht sein. Dabei sollen in erster Linie sämtliche in den drei Exkursionsteilen erhobenen Daten (Entscheidungen und Cue-Rangreihen) sowie die normativ korrekten Lösungen dargestellt werden können. Drei Textfelder die der Erfassung von Anmerkungen zur Testauswahl sollen ebenfalls Platz haben. Dies legte von Anfang an eine Dreiteilung der Oberfläche nahe, wobei in jedem Drittel sämtliche Daten aus jeweils einer der drei Exkursionen dargestellt werden.

Übersichtlichkeit: Trotz des nur beschränkt vorhandenen Platzes konnten wir das Kriterium der Kompaktheit relativ mühelos erfüllen, ohne dass wir bei der Übersichtlichkeit hätten Abstriche machen müssen. Zuunterst auf der Bildschirmseite blieb sogar noch ausreichend Platz um das von der Versuchsperson in den drei Exkursionen erzielte Gesamtergebnis zu vermelden.

7.5.5 Persönliche Angaben

Das Modul zur Erfassung persönlicher Angaben haben wir in dieser Art bereits im „Quizshow-Experiment“ (Christen, Läge & Hausmann, 2007) verwendet. Es handelt sich dabei um ein erprobtes und gut wiederverwendbares Standard-Modul, das vertikal dreigeteilt ist: Die Felder im oberen Drittel dienen der Erfassung von Kontaktinformationen (Wohnadresse, Telefonnummer und Email-Adresse), sowie des Berufs, des Geburtsdatums und des Geschlechts der Versuchsperson. Im mittleren Drittel fragen wir nach der Bereitschaft zur Teilnahme an zukünftigen Experimenten. Durch Anklicken des entsprechenden Radio-Buttons kann uns die Versuchsperson signalisieren, dass sie auch an einem allfälligen Nachfolge-Experiment teilnehmen würde bzw. dass sie für weitere Experimente nicht mehr zur Verfügung stehen wird. Die Angabe eines Kommentars ist optional: Über das Kommentarfeld bieten wir den Versuchsperson die Möglichkeit, ihre Entscheidung auf freiwilliger Basis zu begründen. Die Frage nach der Bereitschaft zur Teilnahme an zukünftigen Experimenten hat einerseits rein rekrutierungstechnische Gründe, kann andererseits aber auch einen Anhaltspunkt für die Interpretation auffälliger Versuchsdaten liefern. So könnte eine fehlende Bereitschaft auf ein ungenügendes commitment hindeuten und entsprechend zum Ausschluss der Versuchsperson bzw. der erhobenen Daten führen, wobei im konkreten Fall natürlich noch weitere Indizien in die Bewertung einfließen bzw. Überlegungen angestellt werden müssen. Im untersten Drittel schliesslich, steht ein grosses Kommentarfeld zur Angabe jeglicher Art von Anmerkungen zum Experiment zur Verfügung.

7.6 Details zur Programmierung

Das Experiment wurde vollständig in der Programmiersprache Microsoft Visual Basic 6.0 implementiert. Vorbereitete Daten (wie bspw. die von den Cues gelieferten Hinweise) waren in einer Microsoft Access Datenbank (Version 2003) hinterlegt. Zur Speicherung der während des Experiments erhobenen Daten wurde dieselbe Datenbank verwendet.

Modularität: Mit Visual Basic erstellte Programme sind nicht streng prozedural aufgebaut, sondern laufen ereignisgesteuert ab. Die zu einer bestimmten Bildschirmoberfläche („form“) gehörenden Ereignisprozeduren sind jeweils in einer eigenen Quellcode-Datei zusammengefasst. Deshalb existieren mindestens so viele Quellcode-Dateien wie es Bildschirmoberflächen gibt. In einer weiteren Datei sind zudem alle nicht ereignisspezifischen Prozeduren und Funktionen untergebracht, wie bspw. die Quicksort-Funktion, die an verschiedenen Stellen im Experiment zur Erzeugung von Zufallsreihenfolgen eingesetzt wird. Neben dieser dateitechnisch aufgezwungenen Modularität, um die man nicht herumkommt, liegt auch der Programmlogik ein modulares Konzept zu Grunde. So sind die forms als programmtechnisch voneinander unabhängige, eigenständige Module implementiert, die beliebig oft und in frei wählbarer Reihenfolge aufgerufen werden können.

Externe Ablaufsteuerung: Im Gegensatz zu früheren Experimenten (vgl. bspw. ...) ist die Steuerung des Ablaufs nicht mehr fest im Quellcode verdrahtet, sondern wird über Einträge in einer speziell dafür eingerichteten Datenbanktabelle gesteuert. Über diese Einträge wird festgelegt, in welcher Reihenfolge und mit welchen Parametern die einzelnen Module aufgerufen werden. Tab. ... gibt einen Überblick über alle in der Datenbank vorhandenen und für das Experiment benötigten Tabellen.

Erweiterbarkeit und Adaptivität: Der modulare Aufbau des Experiments einerseits und die externe, Datenbank-basierte Ablaufsteuerung andererseits sorgen dafür, dass das Experiment relativ leicht erweitert bzw. abgeändert werden kann. So können bereits vorhandene Module entweder zu einem anderen Zeitpunkt aufgerufen oder vollständig aus dem Experiment entfernt werden. Genauso problemlos lassen sich neu erstellte Module in das Experiment einbinden. Zudem können die beim Aufruf eines Moduls übergebenen Parameter sehr einfach adaptiert werden, bspw. um die Anzahl der Runden innerhalb eines Experimentalteils zu erhöhen. Möchte man hingegen die Anzahl der im Information-Board dargestellten Cues ändern, kann dies nicht mehr mittels einfacher Anpassung eines beim Modul-Aufruf übergebenen Parameters bewerkstelligt werden. Das Information-Board müsste in diesem Fall von Anfang auf die Darstellung einer variablen Anzahl Cues ausgerichtet sein. Bei einer Vergrößerung des Cue-Angebots müsste zudem die Datenbank erweitert werden, so dass man auch von den zusätzlich angebotenen Cues Informationen abrufen könnte. Eine inhaltlich relativ geringfügige Änderung am Experiment lässt sich also nicht immer durch die einfache Anpassung eines Parameters bewerkstelligen. Speziell hinsichtlich einer möglichen (nachträglichen) Änderung der Anzahl der verwendeten Cues haben wir deshalb darauf geachtet, dass sich der Umfang der am Quellcode ggf. vorzunehmenden Änderungen in Grenzen hält. Die Benutzeroberfläche müsste aber in jedem Fall angepasst werden (und dies macht i.d.R. den grössten Anteil an der für die Anpassungen benötigten Zeit aus).

Unterbrechbarkeit: Bei der Programmierung des Experiments haben wir darauf geachtet, dass dieses an einer beliebigen Stelle bzw. zu einem beliebigen Zeitpunkt unterbrochen und zu einem späteren Zeitpunkt an derselben Stelle fortgesetzt werden kann – unabhängig davon, ob der Unterbruch beabsichtigt oder unbeabsichtigt (bspw. durch einen Stromausfall) erfolgte. Um dies zu erreichen, werden alle in einem Durchgang erhobenen Daten unmittelbar nach Beendigung des Durchgangs abgespeichert. Zudem wird beim Start des Experiments geprüft, welche Durchgänge schon abgeschlossen bzw. wie viele Daten bereits erhoben worden sind. Im schlimmsten Fall gehen bei einem Unterbruch somit lediglich die Daten aus der Runde verloren, in der das Experiment unterbrochen wurde. In diesem Fall würde das Experiment mit ebendieser Runde fortgesetzt, so dass am Ende für jede Versuchsperson ein vollständiger Satz an erhobenen Daten vorliegt.

Zeiterfassung: Sowohl in der Trainings- als auch in der Anwendungsphase wird die pro Runde benötigte Zeit erfasst und in der Datenbank abgespeichert. Dies ermöglicht im Nachhinein eine minimale Kontrolle des zeitlichen Ablaufs des Experiments.

Dokumentation: Bereits bei der Entwicklung früherer Experimente hat sich das Kommentieren des Quellcodes (mittels direkt im Code angebrachter Kommentare) sehr bewährt. Im vorliegenden Experiment wurde der Quellcode deshalb auf dieselbe Weise kommentiert.

7.7 Auswertemöglichkeiten

Aufgrund der unterschiedlichen Untersuchungsbedingungen werden pro Expedition unterschiedlich viele Cues konsultiert: In Expedition 1 können die Versuchspersonen an einem Äffchen so viele Tests durchführen, bis zum ersten Mal ein Test anspricht. Pro Runde liegen somit 1 bis 5 Cues vor. Aus Expedition 2 liegt immer genau ein Cue pro Runde vor, und in Expedition 3 hängt die Anzahl der vorliegenden Cues davon ab, ob der als erstes konsultierte Cue diskriminiert oder nicht. Falls ja, darf die Versuchsperson auch noch alle anderen Tests an dem Äffchen durchführen, falls nein, muss sie sofort entscheiden, ob sie das untersuchte Äffchen behandeln will oder nicht. Im ersten Fall liegen deshalb neben dem zuerst konsultierten Cue i.d.R. noch 4 weitere Cues vor, deren Reihenfolge dann allerdings willkürlich ist. Dementsprechend beschränken wir unsere Betrachtungen auch in Expedition 3 auf den jeweils ersten konsultierten Cue pro Runde.

Wenn wir nun die Reihenfolgen betrachten in denen eine Versuchsperson in den drei Expeditionen die Tests auswählt, können wir (aufgrund unserer Untersuchungsbedingungen) daraus die subjektiv wahrgenommene Validitätsreihenfolge (Expedition 1), den Cue mit dem höchsten wahrgenommenen Success (Expedition 2) sowie den Cue mit der höchsten wahrgenommenen Diskriminationsrate (Expedition 3) sicher bestimmen.

Auf der Grundlage dieser Daten bieten sich weitere Auswertungen bzw. Kriterien an mit denen wir messen können, wie erfolgreich eine Versuchsperson im Cue-Lernen ist:

1) Abweichung von zufälligem Verhalten (Güte des zuerst konsultierten Cues): Wenn wir wissen wollen ob das Informationssuchverhalten der Versuchsperson (d.h. die Reihenfolge in der sie die

Tests durchgeführt hat) von rein zufälligem Verhalten abweicht, überprüfen wir die Null-Hypothese "Die Vp hat nichts gelernt". Zu diesem Zweck vergleichen wir für jede Expedition mittels einseitigem One-Sample-t-Test den expeditionsabhängigen mittleren Gütewert der fünf Cues (Expedition 1: Validität [V] = .756; Expedition 2: Success [Suc] = .580; Expedition 3: Diskriminationsrate [D] = .468) mit den mittleren Gütewerten der von der Versuchsperson in der Expedition jeweils zuerst durchgeführten Tests.

2) *"Erfüllungs-Score" (ES) pro Expedition und Versuchsperson:* Für jede Aufgabe einer Expedition wird der Gütewert (Validität, Success, Diskriminationsrate) des zuerst durchgeführten Tests in Beziehung gesetzt zum Gütewert des normativ in dieser Expedition anzuwendenden Tests. Bei einer Übereinstimmung mit der normativen Testanwendung (also wenn der erste in der Aufgabe durchgeführten Test genau der ist, der gemäss geltender Norm auch als erster angewendet werden muss), resultiert für diese Aufgabe ein Erfüllungswert von 1. Wird hingegen der normativ schlechteste Test zuerst angewendet, resultiert ein Erfüllungswert von 0. Zwischen diesen beiden Extremen wird ein Messwert linear zur Skala des normativ gerade geltenden Gütewertes berechnet. Der mittlere Erfüllungswert über die 12 Aufgaben ergibt für diese Person in dieser Bedingung den Erfüllungs-Score (ES).

3) *Adaptivität des gezeigten Informationssuchverhaltens auf der Basis der tatsächlichen Cue-Güte:* Eine Versuchsperson könnte in jeder der drei Expeditionen dasselbe Verhaltensmuster zeigen, ihr Verhalten also nicht der entsprechenden Situation anpassen sondern situationsunabhängig immer dieselben Cues bevorzugen. Um etwas über die Adaptivität des Informationssuchverhaltens einer Versuchsperson zu erfahren, berechnen wir deshalb für jede Expedition zusätzlich die beiden Erfüllungs-Scores, die sich ergeben, wenn die beiden jeweils gerade nicht geforderten Normen angewendet werden.

Beispiel: Versuchsperson X bevorzugt in jeder Expedition die validesten Cues. In diesem Fall wären die Erfüllungsscores die sich durch den Transfer der Verhaltensnorm von Expedition 1 (V) auf das Verhalten in Expedition 2 (Success) und Expedition 3 (D) ergeben, grösser als die originalen Erfüllungs-Scores aus den Expeditionen 2 und 3.

4) *Erfüllungs-Scores und Adaptivität auf der Basis der subjektiven Überzeugungen:* Sowohl die Erfüllungs-Scores (Auswertung 2) als auch die Adaptivität des gezeigten Informationssuchverhaltens (Auswertung 3) können analog zu den Auswertungen 2 und 3 auch auf der Basis der im Befragungsteil angegebenen Gütewerte der Cues berechnet werden. Diese Auswertungen tragen jedoch nur dann zum Verständnis des Verhaltens der Versuchspersonen im Experiment bei, wenn die auf der Basis der subjektiven Überzeugungen berechneten Erfüllungs-Scores besser ausfallen als die mit den tatsächlichen Werten berechneten.

5) *Beurteilung des Lernens auf der Basis der subjektiven Überzeugungen:* Auf deklarativer Ebene können zur Beurteilung des Lernens die von den Versuchspersonen im Befragungsteil angegebenen Rangordnungen und Gütewerte mit den tatsächlichen Rängen der Cues (ordinal) und mit den tatsächlichen Ausprägungen der Cuegüte (Pearson-Korrelation) verglichen werden.

Daneben kann für die angegebenen Gütwerte auch eine Abweichung zu den tatsächlichen Werten als Differenzwert berechnet werden.

6) *Anzahl korrekter Entscheidungen pro Expedition*: Für jede Versuchsperson kann die Anzahl korrekter Entscheidungen pro Expedition miteinander verglichen werden. Dabei muss allerdings berücksichtigt werden, dass die maximale Anzahl der erreichbaren korrekten Entscheidungen (bei konsequenter Anwendung der jeweils normativ korrekten Suchstrategie) nicht bei allen 3 Expeditionen gleich gross ist: In Expedition 1 würde die konsequente Anwendung der V-Strategie zu 12 korrekten und 0 falschen Entscheidungen führen. In Expedition 2 würde der Kauf des jeweils ersten Cues pro Aufgabe (bei konsequenter Anwendung der Success-Strategie) 6-mal zu einem korrekten Tipp führen und 1-mal zu einem falschen. Zudem würde man in fünf Fällen keinen Tipp erhalten und raten müssen (Ratewahrscheinlichkeit = 50%). Im Mittel würde man somit lediglich 8 bis 9 korrekte Entscheidungen treffen. In Expedition 3 wiederum würde ein Proband bei konsequenter Anwendung der D-Strategie 10 Mal einen korrekten Tipp erhalten, da der am stärksten diskriminierende Test ($D = .80$) in 10 der 12 Aufgaben anschlägt. In diesen 10 Fällen könnte die Versuchsperson auch die restlichen 4 Cues zu Rate ziehen und bei normativ korrektem Vorgehen (d.h. ihre Entscheidung auf dem validesten Test basierend) 10 Mal eine korrekte Entscheidung treffen. In den verbleibenden 2 Fällen müsste sie raten (Ratewahrscheinlichkeit = 50%) und käme damit im Mittel auf 11 korrekte Entscheidungen.

Die Auswertungen und Resultate des ersten auf dem Äffchen-Szenario basierenden Experiments werden in AKZ-Forschungsbericht Nr. 37 (Christen, Hausmann & Läge, 2006) detailliert beschrieben.

7.8 Anhänge

7.8.1 Rahmengeschichte

Wir haben grossen Wert auf eine kohärente Rahmengeschichte gelegt. Die Versuchspersonen sollen möglichst leicht in das experimentelle Setting „eintauchen“ können und dabei die Perspektive eines Mitglieds eines zoologischen Forschungsteams übernehmen. Hier die Rahmengeschichte im Wortlaut:

Aufgabe und Ziel

Du bist Mitglied eines zoologischen Forschungsteams, das sich in demjenigen Urwaldgebiet befindet, welches die letzten überlebenden Exemplare der vom Aussterben bedrohten Tigerschwanzäffchen beherbergt. In der Nähe dieses Urwaldes befindet sich eine Zuchtstation, die etwa 200 dieser Äffchen beherbergt. Die Grösse der im Urwald frei lebenden Population umfasst gemäss der letzten Zählung noch ca. 50 Tiere.

Leider sind die Tigerschwanzäffchen von einer heimtückischen (möglicherweise genetisch bedingten) Krankheit bedroht. Diese ist äusserlich nicht zu erkennen und führt unbehandelt innert kürzester Zeit zum Tode der betroffenen Individuen. Die Erkrankungsrate der Reservat-Population beträgt ca. 50%. Die Erkrankungsrate der Wild-Population ist nicht genau bekannt.

Erst vor kurzem wurde zufälligerweise ein Mittel (Medikament) entdeckt, welches die von der Krankheit betroffenen Äffchen in jedem Fall rettet, sofern ihnen das Mittel verabreicht wird. Unglücklicherweise sterben die Äffchen aber, wenn sie nicht krank sind, ihnen das Mittel aber trotzdem verabreicht wird.

Das Mittel hat in diesem Fall also die Wirkung eines tödlichen Giftes. Zwei Extremfälle: Würdest Du keinem Äffchen das Mittel verabreichen, so würden alle erkrankten Äffchen (50%) sterben und alle gesunden Äffchen (die anderen 50%) überleben. Im umgekehrten Fall, d.h. wenn Du das Mittel sämtlichen Äffchen verabreichen würdest, würden alle gesunden Äffchen sterben (50%) und alle erkrankten Äffchen (die anderen 50%) überleben. Die Wirkungsweise des Mittels ist bisher nicht bekannt.

Bei der Entscheidung, ob Du einem Äffchen das Mittel verabreichen sollst oder nicht, bist Du glücklicherweise nicht auf raten angewiesen. Es stehen fünf verschiedene Test-Substanzen (Tests) zur Verfügung. Diese Test-Substanzen können einem Äffchen ins Blut gespritzt werden. Dort führen sie dann mit einer bestimmten Wahrscheinlichkeit (Ansprech-Wahrscheinlichkeit) zu einer für den jeweiligen Test charakteristischen Blutbildveränderung. Jeder Test hat seine eigene (konstante) Ansprech-Wahrscheinlichkeit. Die verschiedenen Tests sprechen unterschiedlich gut an, d.h. sie funktionieren unterschiedlich gut. Wenn ein Test tatsächlich anspricht, d.h. wenn er zu einer positiven oder negativen Blutbildveränderung führt, liefert er damit einen Hinweis darauf, ob das untersuchte Äffchen krank ist oder nicht. Diese Hinweise (positive / negative Veränderungen) sind allerdings nicht in jedem Fall korrekt: Jeder der fünf Tests hat seine eigene (konstante) Zuverlässigkeit. Die fünf Tests sind voneinander unabhängig.

Zusammenfassung

- Wenn ein Test nicht anspricht, bleibt das Blutbild gleich wie zuvor (d.h. es kommt zu keiner Blutbildveränderung).
- Wenn ein Test anspricht, bedeutet dies, dass sich das Blutbild entweder positiv oder negativ verändert.
 - ➔ Eine negative Veränderung (-) deutet darauf hin, dass das Äffchen gesund ist!
 - ➔ Eine positive Veränderung (+) deutet darauf hin, dass das Äffchen krank ist!
- Die 5 Tests haben unterschiedliche Ansprech-Wahrscheinlichkeiten und Zuverlässigkeiten.

Deine Aufgabe besteht darin, möglichst viele Äffchen zu retten! Jeweils unmittelbar nach der Untersuchung eines Äffchens (mittels eines oder mehrerer Tests) musst Du Dich für oder gegen die Verabreichung des Mittels entscheiden.

Ablauf

Das Experiment besteht aus 2 Phasen: Einer Lernphase und einer Exkursionsphase:

In der Lernphase wohnst du der Untersuchung der Äffchen aus der Reservat-Population bei. Dabei sollst Du ein Gefühl für die unterschiedlichen Ansprech-Raten und Zuverlässigkeiten der fünf zur Verfügung stehenden Tests entwickeln. Dazu begleitest Du den Veterinärmediziner des Forschungsteams bei der Untersuchung der Äffchen und fällst – basierend auf den Untersuchungsergebnissen – für jedes untersuchte Äffchen einen hypothetischen Behandlungsentscheid (d.h. Verabreichung vs. Nichtverabreichung des Mittels). Unmittelbar nach jedem Behandlungsentscheid erfährst Du, ob das Äffchen die von Dir vorgeschlagene Behandlung (Mittel vs. kein Mittel) überlebt hätte oder nicht.

In der Exkursionsphase wird es dann ernst. Dort wird Deine Aufgabe darin bestehen, möglichst viele Äffchen der Wild-Population richtig zu diagnostizieren und auch tatsächlich zu behandeln (Verabreichung vs. keine Verabreichung des Mittels), so dass möglichst viele kranke Äffchen geheilt werden und möglichst wenige gesunde Äffchen durch die Verabreichung des Mittels verenden.

7.8.2 Instruktionen der drei Expeditionen im Wortlaut

Instruktion für Expedition 1 (Anforderung Validität):

Gemäß Anweisungen der Organisation für die Du arbeitest, musst Du die Untersuchung eines Äffchens aus finanziellen Gründen sofort abbrechen und einen Behandlungsentscheid treffen, sobald ein Test an diesem Äffchen angesprochen hat (d.h. zu einer positiven oder negativen Blutbildveränderung geführt hat). Deine Entscheidung, ein Äffchen zu behandeln oder nicht zu behandeln, basiert somit auf jeweils genau einem einzigen Testergebnis. Die Anzahl an einem Äffchen durchführbarer Tests ist nicht grundsätzlich beschränkt. Du kannst jedoch keine weiteren Tests mehr an dem betreffenden Äffchen durch-

führen, nachdem Du ein erstes (positives oder negatives) Testergebnis für dieses Äffchen erhalten hast.

Wähle deshalb die Reihenfolge der Tests gut aus!

Instruktion für Expedition 2 (Anforderung Success):

Aus finanziellen Gründen verschärft Dein Arbeitgeber für die 2. Expedition die Untersuchungs-Bedingung: Ab sofort ist nur noch genau 1 Test pro Äffchen erlaubt. Unabhängig davon, ob der Test anspricht oder nicht, muss und kann nur genau 1 Test durchgeführt werden. Das Ziel besteht aber nach wie vor darin, möglichst viele Äffchen richtig zu behandeln (d.h. zu retten bzw. nicht umzubringen). Falls also der von Dir gewählte Test nicht ansprechen sollte, kannst Du nur noch raten, ob das Äffchen krank ist oder nicht. Deine Entscheidung, ihm das Mittel zu verabreichen oder nicht, würde damit vom Zufall abhängen. Wähle also diesen einen Test sorgfältig aus!

Instruktion für Expedition 3 (Anforderung Diskrimination):

Für die 3. Expedition lockert Dein Arbeitgeber die Untersuchungs-Bedingung nun wieder ein wenig. Bezüglich der Anzahl erlaubter Tests pro Äffchen gilt nun folgendes: Sofern der erste an einem Äffchen durchgeführte Test anspricht, darfst Du an diesem Äffchen beliebig viele weitere Tests durchführen (da nur in diesem Fall zusätzliche, rechtlich gebundene Spendengelder für die weiteren Tests zur Verfügung stehen). Falls jedoch der erste von Dir gewählte Test nicht anspricht, darfst Du keine weiteren Tests mehr durchführen. Das Überleben des betreffenden Äffchens würde damit von einer Zufallsentscheidung abhängen.

Wähle deshalb auch hier die Reihenfolge der Tests (insbesondere des ersten) sorgfältig aus!

7.8.3 Fragestellungen der Befragungsphase im Wortlaut

1. Relative Häufigkeit korrekter Testergebnisse?

- Angenommen, Du würdest jeden einzelnen Test so oft hintereinander anwenden, bis dass er insgesamt 100-mal angesprochen hat, welcher der 5 Tests würde dann am häufigsten, zweithäufigsten, etc. ein zuverlässiges Ergebnis liefern?
- Angenommen, Du würdest den jeweils nebenstehenden Test (links) so oft anwenden, bis dass er insgesamt 100-mal angesprochen hat, wie oft (von diesen 100-mal) würde er Dir dann ein zuverlässiges Ergebnis liefern?

2. Absolute Häufigkeit korrekter Testergebnisse?

- Angenommen, Du würdest jeden einzelnen Test genau 100-mal hintereinander anwenden, welcher der 5 Tests würde dann am häufigsten, zweithäufigsten, etc. ein zuverlässiges Ergebnis liefern?
- Angenommen, Du würdest den jeweils nebenstehenden Test (links) genau 100-mal hintereinander anwenden, wie oft (d.h. in wie vielen dieser 100 Testanwendungen) würde er Dir dann ein zuverlässiges Ergebnis liefern?

3. Ansprech-Häufigkeit der Tests?

- Angenommen, Du würdest jeden einzelnen Test genau 100-mal hintereinander anwenden, welcher der 5 Tests würde dann am häufigsten, zweithäufigsten, etc. ansprechen?
- Angenommen, Du würdest den jeweils nebenstehenden Test (links) genau 100-mal hintereinander anwenden, wie oft (d.h. in wievielen dieser 100 Testanwendungen) würde er dann überhaupt ansprechen (d.h. eine korrekte oder falsche Vorhersage machen)?

7.8.4 Festgelegte Cue-Informationen

ID	PhaseNr	BlockNr	Cue1	Cue2	Cue3	Cue4	Cue5
1	1	1	0	0	2	0	2
2	1	1	0	2	2	0	2
3	1	1	0	0	0	1	0
4	1	1	0	0	2	0	1
5	1	1	2	0	0	2	2
6	1	1	0	0	2	0	1
7	1	1	0	0	2	1	0
8	1	1	0	2	1	2	1
9	1	1	0	0	0	2	2
10	1	1	0	0	0	2	1
11	1	2	0	2	0	0	2
12	1	2	0	0	0	2	1
13	1	2	0	0	2	1	0
14	1	2	2	0	2	0	2
15	1	2	0	1	0	0	2
16	1	2	0	0	0	2	1
17	1	2	0	0	1	0	0
18	1	2	0	0	2	1	1
19	1	2	0	0	2	2	2
20	1	2	0	2	2	2	1
21	1	3	0	0	0	2	1
22	1	3	2	0	0	1	0
23	1	3	0	2	0	2	2
24	1	3	0	0	2	2	0
25	1	3	0	0	2	0	2
26	1	3	0	0	0	0	2
27	1	3	0	2	1	0	1
28	1	3	0	0	2	1	1
29	1	3	0	0	2	2	2
30	1	3	0	0	2	0	1
31	1	4	0	2	0	2	0
32	1	4	0	0	2	0	1
33	1	4	0	0	2	1	0
34	1	4	0	1	0	0	2
35	1	4	0	0	2	2	2
36	1	4	0	0	2	0	1
37	1	4	2	0	1	2	1
38	1	4	0	0	2	1	2
39	1	4	0	0	0	2	1
40	1	4	0	0	0	0	2
41	1	5	0	0	0	1	1
42	1	5	0	0	2	0	2
43	1	5	2	0	2	2	2
44	1	5	0	2	1	1	2
45	1	5	0	0	2	0	1
46	1	5	0	0	2	1	1
47	1	5	0	0	0	2	0
48	1	5	0	0	2	2	0
49	1	5	0	2	0	0	1
50	1	5	0	0	0	2	2
51	1	6	0	2	0	1	0
52	1	6	2	0	2	0	2
53	1	6	0	0	0	2	1
54	1	6	0	0	1	2	0
55	1	6	0	2	1	2	1
56	1	6	0	0	2	0	2
57	1	6	0	0	2	0	2
58	1	6	0	0	2	0	1
59	1	6	0	0	2	1	2
60	1	6	0	0	0	2	1
61	1	7	0	0	0	2	2
62	1	7	0	0	2	1	1

ID	PhaseNr	BlockNr	Cue1	Cue2	Cue3	Cue4	Cue5
63	1	7	0	2	1	0	1
64	1	7	2	0	2	0	2
65	1	7	0	0	2	1	0
66	1	7	0	0	0	2	2
67	1	7	0	0	2	0	1
68	1	7	0	0	0	2	0
69	1	7	0	1	2	0	1
70	1	7	0	0	0	2	2
71	1	8	2	0	2	1	1
72	1	8	0	0	1	0	0
73	1	8	0	2	0	0	1
74	1	8	0	0	2	2	0
75	1	8	0	0	2	0	2
76	1	8	0	0	0	0	2
77	1	8	0	2	0	1	1
78	1	8	0	0	2	2	2
79	1	8	0	0	2	1	1
80	1	8	0	0	0	2	2
81	1	9	0	2	2	2	0
82	1	9	0	0	2	1	0
83	1	9	0	0	2	0	1
84	1	9	0	0	0	2	1
85	1	9	0	1	0	0	2
86	1	9	0	0	2	2	2
87	1	9	2	0	0	0	1
88	1	9	0	2	1	1	2
89	1	9	0	0	0	2	2
90	1	9	0	0	2	0	1
91	1	10	0	0	2	2	2
92	1	10	0	0	0	2	1
93	1	10	0	0	2	0	0
94	1	10	0	2	2	2	1
95	1	10	0	0	0	0	2
96	1	10	2	0	0	2	2
97	1	10	0	2	1	0	1
98	1	10	0	0	0	1	0
99	1	10	0	0	2	1	1
100	1	10	0	0	2	0	2
101	1	11	0	0	2	0	2
102	1	11	0	2	2	0	2
103	1	11	0	0	0	1	0
104	1	11	0	0	2	0	1
105	1	11	2	0	0	2	2
106	1	11	0	0	2	0	1
107	1	11	0	0	2	1	0
108	1	11	0	2	1	2	1
109	1	11	0	0	0	2	2
110	1	11	0	0	0	2	1
111	1	12	0	2	0	0	2
112	1	12	0	0	0	2	1
113	1	12	0	0	2	1	0
114	1	12	2	0	2	0	2
115	1	12	0	1	0	0	2
116	1	12	0	0	0	2	1
117	1	12	0	0	1	0	0
118	1	12	0	0	2	1	1
119	1	12	0	0	2	2	2
120	1	12	0	2	2	2	1
121	1	13	0	0	0	2	1
122	1	13	2	0	0	1	0
123	1	13	0	2	0	2	2
124	1	13	0	0	2	2	0

7.9 Literatur

- Bröder, A. (2000). Assessing the empirical validity of the "Take The Best"-heuristic as a model of human probabilistic inference. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 26, 1332–1346.
- Christen, S., Hausmann, D. & Läge, D. (2006). *Ist ein Gefühl für Informationsgüte lernbar? Ein Lernexperiment zu Validität, Diskriminationsrate und Success von Informationsquellen*. AKZ-Forschungsbericht 37. Zürich: Angewandte Kognitionspsychologie.
- Christen, S., Läge, D. & Hausmann, D. (2007). *Die Quizshow: Ein experimentelles Design zur Messung von individuellen Informationssuch-Strategien in einfachen Urteilsheuristiken*. AKZ-Forschungsbericht 55. Zürich: Angewandte Kognitionspsychologie.
- Hausmann, D., Christen, S. & Läge, D. (2005). *Und sie lernen es doch! Das Einschätzen der Validität von probabilistischen Hinweiscues und der Einfluss der subjektiven Repräsentation auf das Urteilsverhalten*. AKZ-Forschungsbericht 08. Zürich: Angewandte Kognitionspsychologie.
- Lee, M. D., & Cummins, T. D. R. (2004). Evidence accumulation in decision making: Unifying the “take the best” and the “rational” models. *Psychonomic Bulletin and Review*, 11, 343–352.
- Newell, B. R., & Shanks, D. R. (2003). Take the best or look at the rest? Factors influencing "one-reason" decision-making. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 29, 53–65.
- Newell, B. R., Rakow, T., Weston, N. J., & Shanks, D. R. (2003). Search strategies in decision-making: the success of "success". *Journal of Behavioral Decision Making*, 17, 117–137.
- Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. Cambridge, MA: Cambridge University Press.

Kapitel 8

Ist ein Gefühl für Informationsgüte lernbar? Ein Lernexperiment zu Validität, Diskriminationsrate und Success von Informationsquellen

Entscheidungsrelevante Information ist in vielen Fällen nicht sicher, sondern probabilistisch. Alltagsbeispiele reichen von Börsenempfehlungen bis hin zu medizinischen Tests. Die Güte derartiger Informationsquellen (in Modellen operationalisiert als Validität und Diskriminationsrate) muss in Alltagssituationen in einer konkreten Umwelt durch praktische Erfahrung gelernt werden. Das vorliegende Experiment zum Cue-Probability-Learning stellt diesen erfahrungsabhängigen Lernvorgang für Validitäten und Diskriminationsraten am Beispiel medizinischer Tests nach: In einem Beobachtungssetting sind V und D simultan zu erlernen. Spätere Aufgaben erfordern von den Personen, Tests ganz spezifisch aufgrund ihrer Validität, Diskriminationsrate oder Success-Rate (einer Kombination aus V und D) einzusetzen. Es zeigt sich, dass die Mehrzahl der Personen auf Grund des in der Lernphase erworbenen Wissens in der Lage sind, Validität und Diskriminationsrate als separierte Masse einzusetzen. Etwa die Hälfte der Probanden erlernt zudem die Success-Rate als ein integriertes Mass.

8.1 Einleitung

Mit der „Take The Best“-Heuristik (TTB) haben Gigerenzer, Hoffrage & Kleinbölting (1991) eine einfache, non-kompensatorische und nur auf Validitäten binärer Cues beruhende Heuristik für probabilistisches Entscheiden zwischen zwei Alternativen vorgeschlagen: Information wird in der absteigenden Reihenfolge von Cue-Validitäten betrachtet, und sobald ein Hinweis zwischen den beiden Alternativen diskriminiert, wird diejenige Alternative gewählt, auf die dieser Cue zeigt. Die Konfidenz des Urteils entspricht dann der Validität des diskriminierenden Cues.

Populär wurde diese Heuristik, nachdem Gigerenzer & Goldstein (1996) anhand des systematischen paarweisen Vergleichs aller deutschen Grossstädte im Hinblick auf ihre Einwohnerzahl zeigen konnten, dass man durch die Wahl des besten diskriminierenden Hinweises für die grössere Stadt (z.B. haben Landeshauptstädte, Universitätsstädte, Städte mit Verein in der Fussballbundesliga etc. in der Regel mehr Einwohner als Städte, denen dieses Merkmal jeweils nicht zukommt) genau so gute Entscheidungen treffen kann wie mit einer vollständigen Analyse aller vorliegenden Cue-Informationen. Bröder (2005) gibt einen umfassenden Überblick über die experimentelle Forschung, die durch diese Simulationsstudie angestossen wurde.

In ihrer ursprünglichen Version ist TTB als Heuristik für Informationssuche im Gedächtnis konzipiert. Jedoch lediglich Bröder & Schiffer (2003) haben die Anwendung von TTB als Gedächtnisabrufstrategie experimentell unter die Lupe genommen. Der Grossteil der Untersuchungen konzentriert sich vielmehr auf Informationssuche in der Umwelt, zumal die drei Building Blocks von TTB (Suchregel, Stopppregel und Entscheidungsregel) hier genau so gut anwendbar und experimentell wesentlich leichter zu beobachten sind. Da Informationssuche in der Umwelt jedoch realiter mit Aufwand verbunden ist, muss für die Suchregel neben der Validität (V) auch die Diskriminationsrate (D) ins Kalkül gezogen werden. Newell, Rakow, Weston & Shanks (2003) sowie Läge, Hausmann, Christen & Daub (2005) berücksichtigten dies in ihren Experimenten: Sie zogen mit Usefulness ($V \times D$) und mit Success ($V \times D + \text{guess} \times (1-D)$) zwei Kombinationen von Validität und Diskriminationsrate ins Kalkül. Usefulness als multiplikative Kombination gibt die Prozentzahl an, in der ein Cue eine korrekte Vorhersage macht (dazu muss er erst einmal diskriminieren und dann richtig vorhersagen). Success berechnet die Prozentzahl, in der man mit Zuhilfenahme dieses Cues eine korrekte Vorhersage macht, wenn man im nicht-diskriminierenden Fall rät. Damit ist Success normativ der beste Cue, wenn man sich nur genau einen Cue anschauen will oder kann (in diesem Fall gilt als Formel: Usefulness für den Fall der Diskrimination plus Ratequote für den Fall der Nicht-Diskrimination). Usefulness hingegen ist eher ein relatives Mass, das weder die Alternativenzahl (durch die Ratequote) berücksichtigt noch für sich genommen die Cuegüte im gesamten Ausprägungsbereich perfekt beschreibt. (Beispielsweise erzielt eine Münze als Cue im Zwei-Alternativen-Fall die enorm hohe Usefulness von .50: sie diskriminiert immer, hat jedoch in ihrer Vorhersagekraft nur Ratequote. Ein perfekter Cue mit $D < .50$ hingegen wäre im Vergleich weniger „useful“).

Success bzw. Usefulness werden also nur dann relevant, wenn limitierte Informationssuche betrieben werden soll. In diesem Fall ist es eine notwendige Voraussetzung für gutes Entscheiden,

dass Menschen entweder ein aus V und D kombiniertes Gütemass für einen Cue direkt durch Beobachtung lernen können oder dass sie V und D einzeln lernen und daraus ein kombiniertes Gütemass durch Kalkulation inferieren. Um TTB (als gedächtnisbasierte Heuristik) einsetzen zu können, müssen Menschen hingegen „nur“ über die Fähigkeit verfügen, durch Beobachtung ihrer Umwelt die Validitäten von Cues akkurat einschätzen zu können. Selbiges gilt auch für eine in den Building Blocks ganz an TTB angelehnte Heuristik mit aktiver Informationssuche, bei der die Cues grundsätzlich diskriminieren (Hausmann, Christen & Läge, 2005, stellen mit dem Quiz-Design die Struktur einer solchen Umwelt her; s.u.).

Diese basale Anwendungsvoraussetzung teilen sich non-kompensatorische Heuristiken mit den kompensatorischen Modellen, die die Psychologie des probabilistischen Entscheidens durchziehen: Auch hier ist es für eine gute Entscheidung notwendig, ein akkurates Bild über die Validitäten von Hinweisinformationen zu besitzen: sonst kann man sie lediglich im Sinne eines ungewichteten Tallying (s. Dawes, 1979) integrieren. Im Kontext kompensatorischer Entscheidungsregeln ist denn auch mehrfach gezeigt worden, dass Menschen fähig sind, solche relativen Häufigkeiten ziemlich akkurat einzuschätzen (Attneave, 1953; Shuford, 1961; Peterson & Beach, 1967; Estes, 1976; Hasher & Zacks, 1979; 1984. Zum Teil werden Einschränkungen berichtet, z.B. merkt Estes an, dass Cues in der Lernphase gleich häufig vorkommen müssen, um Verzerrungen zu vermeiden).

Mit diesen – allerdings aus sehr heterogenen Forschungsfeldern herangezogenen – Befunden im Rücken könnte man sich zufrieden geben und fortan in der experimentellen Forschung wie in der Theoriebildung³⁹ so tun, als sei das Lernen der Validitäten lediglich eine selbstverständliche Formalität. In diesem Fall würde man den Versuchspersonen die Validitäten, Diskriminationsraten und den Success so weit als für die jeweilige Fragestellung nötig vorgeben und dies schlicht und einfach als experimentelle Abkürzung betrachten. Bröder (2000, 2005) argumentiert beispielsweise in diesem Muster, und auch die Zürcher Forschergruppe um Läge und Hausmann gibt ihren Versuchspersonen in der Regel die erforderlichen Kennwerte für die Güte der Cues vor.

Auf der anderen Seite liegen jedoch einige Experimente vor (in erster Linie solche aus der Londoner Gruppe um Newell & Shanks), die sich explizit auf TTB beziehen und einer Anwendungsphase eine Lernphase der Cuegüte voranstellen. Also sollten die vorliegenden Experimente einmal aus der Warte des Erlernens der Cue-Qualitäten (Validität und, soweit erfolgt, Diskriminationsrate und Success) betrachtet werden. Es muss allerdings betont werden, dass das Lernen bis auf das Experiment von Hausmann, Christen & Läge (2005) nie im Mittelpunkt der Fragestellung stand und deswegen für die jeweiligen Autoren nur Randbefunde darstellte. Deswegen werden wir im Folgenden die Resultate zum Cue-Lernen etwas anders und in weiten Passagen auch ausführlicher kommentieren als die Autoren selbst dies taten.

³⁹ Gigerenzer et. al. (1991) machen in der PMM-Theorie denn auch explizite Aussagen über die Repräsentation von Cue-Validitäten im Indikativ: „... cue validities are learned by observing the frequencies of co-occurrences in an environment“ (p. 510).

Bröder (2000) schaltete den Experimenten 1 und 2 jeweils eine umfassende Lernphase vor, in denen die Versuchspersonen die Validitäten binärer Cues für den Kulturstand extraterrestrischer Völker lernen konnten. Sie hatten dabei jeweils 50 Durchgänge für einen einzelnen Cue. Zusätzlich erhielten die Versuchspersonen mit einem graphischen Balken die relative Häufigkeit der korrekten und der falschen Vorhersagen des gerade zu lernenden Cues angezeigt. Mit dieser visuellen Unterstützung hatten es die Versuchspersonen leicht, ein akkurates Gefühl für die vier Cue-Qualitäten aufzubauen. Die Frage nach dem Erlernen der Validität beantwortet dieses Vorgehen aber nicht. Für die beiden weiteren Experimente dieser Studie (sowie für mehr als ein Dutzend Experimente, die Bröder danach zu TTB durchführte) argumentierte der Autor, dass sich das erfolgreiche Erlernen der Validitäten als für derartige Experimente zu aufwändig erwiesen hätte, weshalb man den Probanden die Validitätsordnung der Cues auch direkt vorgeben könne.

Newell & Shanks (2003) übernahmen das Börsenszenario von Bröder (2000, Experiment 3 und 4) und adaptierten es für ihre Fragestellungen. In diesem Szenario sind in Form einer einfachen Tabelle vier binäre Cues vorhanden, um in unverbundenen Einzeldurchgängen von jeweils zwei Aktien die erfolgversprechendere auswählen zu können. Die Cues können (gegen Informationskosten) sequentiell in beliebiger Reihenfolge betrachtet werden. Während Bröder in diesem Szenario die Cues stets in absteigender Validitätsfolge untereinander präsentierte (und seine Versuchspersonen darüber informiert hatte), mussten die Probanden in den Experimenten von Newell & Shanks die vier Validitäten zunächst lernen.

In Experiment 1 waren die Validitäten auf .80, .75, .70 und .69 festgesetzt (und die Diskriminationsraten konstant auf 0.50, so dass sie keine Rolle spielten). Die Validitäten lagen also ziemlich dicht beieinander. Die Versuchspersonen hatten 60 Trainingsdurchgänge, in denen ihnen alle acht Cue-Informationen simultan gezeigt wurden. Sie gaben auf dieser Basis ihren Tipp ab und erhielten daraufhin ein Feedback, welche Aktie die tatsächlich gewinnende sei. Ausserdem wurde ihnen die exakte bayesianisch berechnete Wahrscheinlichkeit dafür angegeben, mit der die von ihnen gewählte Alternative die bessere gewesen wäre.

Für die Betrachtung des Lernerfolgs standen zwei Masse zur Verfügung: aus 120 Testaufgaben mit sequentiellem Cue-Kauf die Häufigkeit, mit der ein Cue gekauft wurde (von den Autoren als „on-line ranking“ bezeichnet), sowie numerische Angaben über die empfundene Nützlichkeit der Cues (erhoben in einer Befragung am Schluss des Experimentes). Mit den Resultaten kamen die Autoren zu dem Schluss, dass die Validitätsfolge nur unzureichend gelernt worden war. In der Tat wählten lediglich 6 der 16 Probanden den validesten Cue am häufigsten, und es waren auch nur diese sechs Personen, die den validesten Cue als den nützlichsten einschätzten. Genau so viele Personen hingegen präferierten einen der beiden Cues mit der niedrigsten Validität. Nur drei Personen repräsentierten in den Häufigkeiten der gekauften Cues die korrekte Validitätsfolge, nur eine einzige Person in der Nachbefragung. Lediglich bei 9 der 16 Personen stimmte überhaupt der als am nützlichsten angegebene Cue mit dem meistgekauften überein.

Über die Reihenfolge, in der die Cues gekauft wurden, berichten die Autoren nichts. Die über die gesamte Stichprobe gemittelten Häufigkeiten der Cue-Käufe zeigen zwar nach dem Jonckheere-Test einen auf dem 5%-Niveau signifikanten Trend in Richtung Validitätsfolge, jedoch wurde der zweitvalideste Cue etwas häufiger gekauft als der valideste und der viertvalideste deutlich häufiger als der drittvalideste. Auch in der eingeschätzten Nützlichkeit rangiert der zweitvalideste Cue knapp vor dem validesten.

Diese Resultate veranlassten die Autoren, für Experiment 2 und 3 der Studie den Versuchspersonen die korrekten Validitätsfolgen jeweils direkt mitzuteilen. Sie taten dies in Form eines „hint“, den sie in die Mitte der Lernphase einstreuten und direkt vor Beginn der Testphase noch einmal wiederholten. Auch Newell, Weston & Shanks (2003) schalteten ihren Experimenten (Börsenszenario mit sechs bzw. zwei Cues) eine Lernphase (von diesmal sogar 180 Durchgängen) vor. Genauso wie Newell & Shanks (2003) gaben sie den Versuchspersonen jedoch zweimal einen „hint“, d.h. eine korrekte Liste der Validitätsfolgen. Dies geschah nach 60 und nach 120 Lerndurchgängen. Auf diese Weise wird es unmöglich zu sagen, was die Versuchspersonen in diesen Experimenten nun wirklich durch Beobachtung gelernt und was sie sich aus den Vorgaben gemerkt hatten. Bis auf jeweils ganz wenige Versuchspersonen folgten alle (zumindest ungefähr) in der Suchreihenfolge dem Validitätspattern. Doch wegen der „Hints“ sagt dies nichts darüber aus, ob und in welchem Umfang die 180 Lerndurchgänge zum Validitätswissen beigetragen haben.

Ohne derartige direkte Hinweise kam das Experiment von Lee & Cummins (2004) aus. Die Versuchspersonen erhielten hier jeweils Paare aus einer Gesamtmenge von 16 Objekten zusammen mit sechs binären Cue-Angaben zu jeder der beiden Alternativen. Sie entschieden sich auf der Basis der Cues für eine Alternative und erhielten unmittelbares Feedback, welche Antwort die richtige war. Auf diese Weise sollten die Probanden ein Gespür für die Höhe der Validitäten der sechs eingesetzten Cues bekommen. Diese Lernphase war die dem eigentlichen Experiment (bestehend aus fünf Aufgaben, bei denen TTB und weighted-additive-Strategie gemäss der Verteilung der zu lernenden Validitäten unterschiedliche Vorhersagen machten) vorgeschaltet. Gleich im ersten von sechs Lernblöcken (19 Paare), als die Versuchspersonen also noch keine Chance gehabt hatten, viel über die Validitäten zu lernen, erzielten sie aufgrund des Vergleichs der Cue-Ausprägungen bereits 72% Treffer. Im zweiten Durchgang (20 Paare) stieg dieser Wert auf 79% an, ein deutliches Signal, dass in irgendeiner Form eine Verbesserung stattgefunden hat. Nach weiteren vier Durchgängen zu je 20 Paaren liegt die Trefferquote schliesslich bei 81%.

Die Autoren schlussfolgern daraus, dass die Versuchspersonen in ihrem Experiment zumindest einige der Cue-Validitäten schnell gelernt hätten (sie verweisen darauf, dass bereits mit dem ersten Durchgang die Quote deutlich über dem Ratelevel von 50% liege und dass sich der Lernerfolg der Marke von 86% annähere, die in ihrem Datenset der Vorhersagegüte sowohl von TTB als auch der weighted-additive-Strategie entspreche, also bei perfektem Wissen der Validitäten zu erzielen sei). Was die Autoren nicht erwähnen ist, dass ein einfaches Tallying (also ohne irgendwelches Wissen über die Validitäten) bereits in 91 der 119 getesteten Fälle (76,5%) zum korrekten Ergebnis und nur in 12 Fällen (10,1%) in die Irre führt. In den 16 Fällen, in denen beide Alternativen gleich viele positive Cue-Ausprägungen haben (13,4%), liegt man durch Raten

auch noch einmal in acht Fällen richtig (6,7%), so dass sich ohne jegliches Wissen über die Cue-Validitäten allein mit Auszählen der positiven Cues ein Erwartungswert von 83,2% ergibt. Selbst den erreichen die Probanden nicht einmal am Ende der Lernphase, so dass sich aus dieser Lernphase heraus kein Ansatzpunkt ergibt zu schlussfolgern, die Versuchspersonen hätten die Validitäten erlernt.⁴⁰ So interessant der weitere Aufbau des Papers auch ist, über die Erlernbarkeit der Validitäten lässt sich auch durch das von Lee & Cummins präsentierte Experiment keine klare Aussage treffen.

Es ist an dieser Stelle noch einmal zu betonen, dass die bislang präsentierten Experimente das Erlernen der Validitäten jeweils nur als Hilfsmittel für die Untersuchung der eigentlichen Fragestellungen verwendet haben. Gerade die Arbeit von Newell & Shanks (2003) konnte jedoch zeigen, dass das Erlernen der Validitäten sehr problematisch sein kann. Entsprechend betonen die Autoren denn auch den grossen Unterschied zwischen ihren Lerndurchgängen und den (ihrer Meinung nach zu TTB-favorabel gestalteten) Lerndurchgängen von Bröder (2000; Experiment 1 und 2). Die „Hints“, die sie danach in ihren Experimenten einbauen, machen die Lernphase denn auch praktisch obsolet, so dass sie letztlich denselben Weg einschlagen wie Bröder, der die Versuchspersonen ebenfalls direkt über die Rangordnung informiert. Für die jeweils bearbeiteten Fragestellungen ist das ohne Zweifel ein probates Vorgehen, denn man möchte ja gerne, dass sich möglichst alle Probanden einer Stichprobe beim Entscheiden auf dieselbe Cue-Ordnung beziehen.

Stärker in den Mittelpunkt rückten Hausmann, Christen & Läge (2005) das Lernen der Validitäten. Sie führten ihre Versuchspersonen in ein Quiz mit schwierigen Wissensaufgaben ein, in dem es jeweils vier Lösungsalternativen gab. In einer ersten Trainingsphase (50 Durchgänge) mussten die Versuchspersonen die richtige Antwort ohne fremde Hilfe finden. Nach dem Eintippen der gewussten oder erratenen Alternative erhielten sie die korrekte Lösung angezeigt sowie eine ausführliche Erklärung (ein mehrzeiliger Text). Zusätzlich wurde den Versuchspersonen das jeweilige Antwortverhalten (Tipp) von vier „Personen“ eingeblendet, welches jeweils mit der korrekten Lösung verglichen werden konnte. Durch das Beobachten des Antwortverhaltens der vier „personalisierten“, Cues konnte auf deren Güte (Validität) geschlossen werden. Im eigentlichen Testdurchgang (20 weitere Fragen) hatten die Versuchspersonen die Möglichkeit – gegen eine Gebühr –, einen oder mehrere der Personen aktiv und sequentiell um deren Tipp zu fragen.

⁴⁰ Auch aus dem Verhalten in den fünf Items des Testdurchgangs lassen sich nur sehr bedingte Schlussfolgerungen in Bezug auf das Validitätswissen ziehen. Immerhin liessen sich diese Items nicht mit Tallying lösen (jeweils gleich viele positive Cueausprägungen für beide Alternativen). In allen fünf Aufgaben kam jedoch der valideste Cue als diskriminierender Cue vor. (Im Trainingsdurchgang zeigte er in 59 der 60 diskriminierenden Fälle in die korrekte Richtung). Fünf der 40 Versuchspersonen orientierten sich in den fünf Testaufgaben an ihm. Möglicherweise hatten sie erkannt, dass dieser Cue praktisch deterministisch ist. (In Frage gestellt wird diese Lernleistung jedoch durch den Umstand, dass keine einzige dieser Personen auf den jeweils mit der Antwort verbundenen Konfidenz-Ratings auf der fünfstufigen Skala konsequent Werte von 4 oder 5 angab. Das sollte aber der Fall sein, wenn ein Cue mit einer Validität von >.98 identifiziert worden ist). Die meisten der Probanden (21 der 40) zeigten ein inkonsistentes Verhalten, welches mit einer Zufallsverteilung von Antwortmustern beschreibbar ist. 14 Probanden entschieden sich konsequent gegen den validesten Cue.

Das Experiment war in zwei Validitätsbedingungen implementiert: Während der höchste Cue ($V_1 = .92$) und der niedrigste Cue ($V_4 = .52$) jeweils identisch waren, wurden die Validitäten des zweiten und dritten Cues in Bedingung A mit $V_2 = .76$ und $V_3 = .60$ weiter von V_1 entfernt als in Bedingung B ($V_2 = .84$ und $V_3 = .68$). Hinsichtlich des besten Cues sollte sich Bedingung A damit als leichter erweisen als Bedingung B. Die vier Auskunftspersonen gaben grundsätzlich zu jeder Frage eine Antwort; die Diskriminationsrate lag somit konstant bei 1.00. Die kniffligen Fragen, das einbringbare Wissen, abzugebende Konfidenzratings sowie nicht zuletzt die ausführlichen Erläuterungen zu den Antworten sorgten dafür, dass die Versuchspersonen nicht einfach im Kopf die Anzahl richtiger Lösungen der vier späteren Auskunftspersonen mitzählen konnten.

In der sich anschliessenden Testphase (20 Durchgänge) gaben die Probanden jeweils wiederum zuerst ihren eigenen Tipp ab und konnten dann (gegen eine hohe Gebühr von jeweils 40% des möglichen Gewinns) noch eine oder mehrere Auskunftspersonen befragen. Danach konnten sie ihre Antworten revidieren oder bei ihrer ursprünglichen Einschätzung bleiben. Neben dieser Verhaltensbeobachtung (Häufigkeit, mit der jede der vier Auskunftspersonen konsultiert wurde) lag auch noch eine direkte Befragung nach der eingeschätzten Validität vor. Sowohl die Versuchspersonen aus Bedingung A als auch diejenigen aus Bedingung B schätzten die Rangordnungen der Validitäten im Mittel korrekt ein, jedoch wurden jeweils die vier präsentierten Cues in den tatsächlichen Ausprägungen massiv unterschätzt (um durchschnittlich knapp 16 Prozentpunkte). Zehn der insgesamt 32 Versuchspersonen schätzte die Validitätsrangfolge vollständig korrekt ein, bei 70% der Probanden war die Rangkorrelation zwischen geschätzten Validitäten und tatsächlichen Validitäten > 0.80 (d.h. maximal eine Rangplatzvertauschung). Allerdings schätzten 11 Versuchspersonen (immerhin ein Drittel der Stichprobe) den tatsächlich zweitvalidesten Cue als den validesten ein, und acht weitere Personen machten keinen klaren Unterschied zwischen dem validesten und zweitvalidesten Cue. Die beiden Bedingungen A und B unterschieden sich übrigens nicht signifikant voneinander, d.h. die Grösse des Unterschieds zwischen dem validesten und zweitvalidesten Cue hatte keinen Einfluss auf die Resultate.

Diese Ergebnisse fallen also deutlich positiver aus als die von Newell & Shanks (2003). Allerdings liegen die Validitäten auch wesentlich weiter auseinander, was die Unterscheidung massiv erleichtern sollte. Auf der anderen Seite vermittelt das Quiz-Design deutlich mehr Ablenkung, und es fehlen die Hinweise auf die exakten Wahrscheinlichkeiten, die Newell & Shanks ihren Versuchspersonen nach jeder Übergangsaufgabe präsentierten. Weitere Stolpersteine im direkten Vergleich der Resultate bestehen darin, dass Hausmann et al. ein Vieralternativen-Design wählten (bei dem die Validitäten demnach zwischen .25 und 1.00 schwanken konnten, dass sie dafür aber die Diskriminationsraten auf 1.00 setzten (was ein effizienteres Beobachten der Trefferraten der Auskunftspersonen ermöglichte).

Methodologisch ist an der Arbeit von Hausmann et. al. zu kritisieren, dass sie die Ratings über die Validitäten in die Lernphase einstreuten, indem sie nach jeweils 10 Durchgängen eine Abfrage des momentanen Eindrucks machten. Es blieb jedoch in den Antwortmustern unklar, ob sich die Versuchspersonen mit ihren Validitätseinschätzungen auf alle bisherigen Durchgänge oder nur auf die jeweils letzten zehn Aufgaben bezogen. Die Autoren entschieden sich nach detaillier-

ter Begutachtung der individuellen Verläufe der Ratings dazu, die Werte über die fünf Abfragen zu mitteln und sich nicht allein auf die letzte Abfrage zu verlassen. Die Schlussfolgerung, dass die Validitätsordnung im Wesentlichen gut gelernt wurde, kann also nur unter Vorbehalt gezogen werden und bedarf weiterer experimenteller Bestätigung.

Die bislang präsentierten Arbeiten beschränken das Lernen der Cue-Güte allein auf die Validität. Newell, Rakow, Weston & Shanks (2003) argumentieren, dass das Lernen von Cue-Güte im Alltag eigentlich gar nicht zum Wissen über Validitäten führt, sondern wegen der steten Konfundierung mit der Diskriminationsrate gleich zum integrierten Konzept des Success. Die Autoren führten deswegen zwei Experimente durch (Börsendesign wie Newell & Shanks, 2003), in denen zusätzlich zur Validität auch die Diskriminationsrate variierte. Sie schalteten den Testphasen jeweils eine Lernphase vor, in der vier Cues mit ihren binären Ausprägungen gleichzeitig präsentiert wurden. Ausserdem wurde den Probanden – wie bereits bei Newell & Shanks (2003) – nach jedem Tipp die exakte bayesianisch berechnete Wahrscheinlichkeit angegeben, mit der die gewählte Alternative die bessere sei. (Dies ist ein Hinweis, der die von den Autoren eingeforderte Alltagsnähe wieder fraglich macht. Denn in der Regel können Menschen nur den tatsächlichen Ausgang in Verbindung mit vorliegenden Cues beobachten, aber nicht die exakten, aufgrund der Validitäten bereits perfekt berechneten Wahrscheinlichkeiten.)

Die Versuchspersonen hatten innerhalb von 64 Durchgängen die Gelegenheit, die Güte der Cues zu erlernen (und zwar nach Validität und Diskriminationsraten, und somit auch nach Success bzw. Usefulness als integriertem Mass). In Experiment 1 stimmte ihr späteres Informationssuch-Muster im Mittel mit der Success-Reihenfolge der vier Cues überein. Auch in einer Nachbefragung darüber, wie „useful“ die Cues seien, entsprach die Ordnung der Nennungen auf Stichprobenebene der Success-Ordnung.

Über die individuellen Lernerfolge erfährt man nur so viel, als dass der paarweise Vergleich zwischen den individuellen Kaufhäufigkeiten lediglich zwischen dem Cue mit höchstem und dem mit niedrigstem Success signifikant wurde. Ausserdem schauten sich 9 der 20 Probanden zuerst den Cue mit dem höchsten Success (und gleichzeitig der höchsten Diskriminationsrate) an und drei Personen den Cue mit der höchsten Validität. Die Autoren lassen es unkommentiert, dass 40% ihrer Versuchspersonen damit einen Cue bevorzugten, der bei korrektem Lernen der Cuegüte eigentlich nach keiner der drei Strategien der Spitzen-Cue ist. Dieser Befund ist nur damit zu erklären, dass ein substantieller Teil der Versuchspersonen den – gleich nach welcher Strategie – besten Cue nicht gelernt hat. Von den vier Cues entsprechen zwei den drei möglichen Strategien, zwei nicht. Hätte keine der Versuchspersonen etwas gelernt, so sollten 10 zufällig einen Cue bevorzugen, der einer der Strategien entspricht. Gefunden wurden 12 Personen, was in einem Binomialtest gegen die Wahrscheinlichkeit von 50% nicht signifikant würde. Aus den berichteten individuellen Daten kann also nicht geschlossen werden, dass die Cue-Qualitäten systematisch gelernt wurden.

Ein Vergleich der nachträglichen Befragung der Nützlichkeit mit der Häufigkeit der Cue-Käufe ergab, dass nur 12 der 20 Personen die gleiche Reihenfolge angaben. Vier Personen verwendeten den als am nützlichsten deklarierten Cue nicht am häufigsten.

In Experiment 2 wurde den Versuchspersonen zusätzlich nach dem Experiment Validität, Diskriminationsrate und Success erklärt. Sie sollten dann eine Rangordnung der vier Cues jeweils gemäss dieser Gütekriterien bilden. Das gelang den Versuchspersonen im Mittel für die Diskriminationsrate und für Success, für Validität gaben sie jedoch dasselbe an wie für Success. Aufgrund dieser Mittelwerte muss geschlossen werden, dass die Versuchspersonen die Validität nicht gelernt haben.

Gleichzeitig wurden bei dieser Befragung die Validitäten massiv unterschätzt (im Mittel um .16), während die Diskriminationsraten systematisch überschätzt wurden (im Mittel um .09). Die Usefulness-Ratings entsprechen in den Grössenordnungen ungefähr der mit $V \times D$ berechneten Usefulness, entsprechen dem aber nicht in der Rangordnung, sondern repräsentieren hier Success.

An Auswertungen auf individuellem Level berichten die Autoren, dass 13 der 24 Probanden einen Cue am häufigsten wählten, der dem besten Cue nach Validität, Success oder Diskriminationsrate entspricht. 11 Personen (46%) wählten also meistens einen Cue, der sich mit keinem dieser drei Modelle als favorabler Cue erweist. Genau so wie in Experiment 1 kann also die Nullhypothese nicht verworfen werden, dass die Personen die Cue-Qualitäten nicht akkurat gelernt haben. Anders als noch in Experiment 1 war im Vergleich der Cue-Präferenzen über alle Cue-Paare nicht einmal der Unterschied zwischen dem häufigsten präferierten und dem am wenigsten häufig präferierten signifikant. Immerhin fanden die Autoren hinsichtlich der nachträglich erhobenen Ratings einen signifikanten Trend (nach Page-Test) zwischen dem als am höchsten und dem als am wenigsten diskriminierenden Cue eingeschätzten. Und für Success ergab sich ein signifikanter Trend zwischen den beiden erfolgreichsten und dem am wenigsten erfolgreich eingeschätzten Cue.

Dieses sind die beiden einzigen bislang vorliegenden Experimente, in denen die Versuchspersonen Validitäten und Diskriminationsraten zugleich zu lernen hatten. Die Resultate sind wenig ermutigend, denn die berichteten Daten deuten nicht darauf hin, dass die Probanden substantiellen Lernerfolg aufweisen konnten. Als positives Ergebnis bleibt festzuhalten – und darauf kommt es Newell et al. auch an –, dass Success deutlich häufiger angewendet und auch akkurater eingeschätzt wurde als die Validität.

Versuchen wir, eine Summe aus diesen sehr heterogenen Befunden zu ziehen: Die Resultate zum Erlernen der Validitätsordnung variieren zwischen „gut“ (Hausmann et al.) und „katastrophal“ (Newell et al.). Wo numerische Einschätzungen der Validitäten stattfanden, wurden diese substantiell unterschätzt (interessanterweise jeweils um durchschnittlich .16). Success scheint aus dem Beobachten heraus weniger schwer zu lernen zu sein, jedoch fallen die Resultate auf Personenebene alles andere als überzeugend aus (genau genommen kann von einem mehrheitlichen Lernerfolg eigentlich keine Rede sein). Die Diskriminationsrate ist (in der einzigen vorhandenen

Nachbefragung) recht ordentlich repräsentiert, jedoch spielt dies keine Rolle, da Success – wenn überhaupt – bereits von vornherein als eigenständiges Gütemass gelernt worden sein muss.

Das Fazit hinsichtlich der Lernbarkeit fällt also nicht sonderlich enthusiastisch aus. Sollte es jedoch so sein, dass Validität noch schlechter gelernt wird als Success, dann würde dies nicht nur TTB den Boden unter den Füßen wegziehen, sondern auch allen auf gewichteter Informationsintegration basierenden Modellen. Es blieben dann nur noch non-kompensatorische Heuristiken mit aktiver Informationssuche (die brauchen Success, aber nicht Validität) und ein ungewichtetes Abzählen der Richtungen der diskriminierenden Cues (ganz im Sinne der Dawes' Rule).

Jedoch erscheinen uns die bisherigen Experimente zu wenig auf das Erlernen der Cue-Güte hin fokussiert zu sein, um ein solches Fazit bereits ziehen zu können. Wir legen deswegen hiermit ein Experiment vor, welches sich einzig und allein dieser Frage widmet und in seiner Konstruktion keine Kompromisse an andere, simultan zu beantwortende Fragestellungen macht.

Dabei soll die Lernsituation so gestaltet sein, dass sowohl Validität als auch Diskriminationsrate als auch Success bzw. Usefulness durch aktives Beobachten erlernbar sind. Es sollen den Versuchspersonen keine Hinweise auf eine bestimmte Cue-Ordnung induziert werden, noch sollen sie graphische Hilfen oder eine vorauslaufende Erklärung der Cue-Güte-Konzepte erhalten, auf die sie zu achten hätten. Allerdings sollen sie durchaus wissen, dass sie später im Experiment die Güte der Cues dringend benötigen werden.

Die Anwendungsphase sollte nicht ein „forced choice“ zwischen womöglich gelernten Cue-Rangordnungen erzwingen (so wie das im Rahmen der Hauptfragestellung von Newell et al. durchaus noch sinnvoll war). Vielmehr sollten die Versuchspersonen die Möglichkeiten bekommen, getrennt voneinander auf Verhaltensebene zu demonstrieren, ob sie Validitäten, Diskriminationsraten und Success gelernt haben.

Wie weiter oben erläutert, ist die Success-Rate das normativ entscheidende kombinierte Mass aus Validität und Diskriminationsrate. Usefulness hingegen ist leichter als Konzept zu erfragen, und die Antworten der Versuchspersonen von Newell et al. auf die Frage „how useful“ tendierten auch numerisch in Richtung Usefulness (Angaben zwischen .20 und .30 werden berichtet) und stellten nicht eine Schätzung des Success dar (Success-Angaben müssten grundsätzlich $>.50$ ausfallen). Aus diesem Grund wählt das im Folgenden berichtete Experiment genau wie die Arbeit von Newell et al. Success als zu überprüfende Verhaltensnorm und Usefulness als Konzept der expliziten Befragung nach einem kombinierten Mass aus Validität und Diskriminationsrate. Es ist in der Konstruktion der Cue-Qualitäten darauf zu achten, dass in der Verhaltensnorm Success und Usefulness auf denselben besten Cue hinweisen. Vom am höchsten diskriminierenden Cue hingegen soll sich der beste Success-Cue – anders als bei Newell et al. – deutlich unterscheiden.

Auf diese Weise sollte es möglich sein, für jede einzelne Versuchspersonen zu bestimmen – und zwar sowohl auf Verhaltensebene als auch auf der Ebene deklarativen Wissens –, welche der drei Cue-Gütemasse sie gelernt hat.

Auf der Basis der Argumentation von Newell et al. wäre dann zu erwarten, dass es eine substantielle Anzahl von Personen gibt, die zwar Success beherrschen, nicht jedoch sowohl Validität als auch Diskriminationsrate. Diese Personen hätten dann ein Gefühl für Success erworben, welches sich nicht aus den separat gelernten Komponenten zusammensetzt, sondern als solches durch Beobachtung gelernt wird. Sollte die TTB-Suchregel (absteigende Ordnung der Validitäten) Bestand haben, so müsste Validität von den Versuchspersonen gut zu lernen sein (unabhängig davon, was mit den anderen Gütemassen geschieht, denn die sind ja für TTB nicht relevant). Für non-kompensatorische Urteilsheuristiken reicht es aus, wenn die lexikographische Rangordnung der Cue-Güte gelernt wird. Für kompensatorische, gewichtete Informationsintegration hingegen müssen die Versuchspersonen zwingend ein Gefühl für die absoluten Validitäten entwickeln.

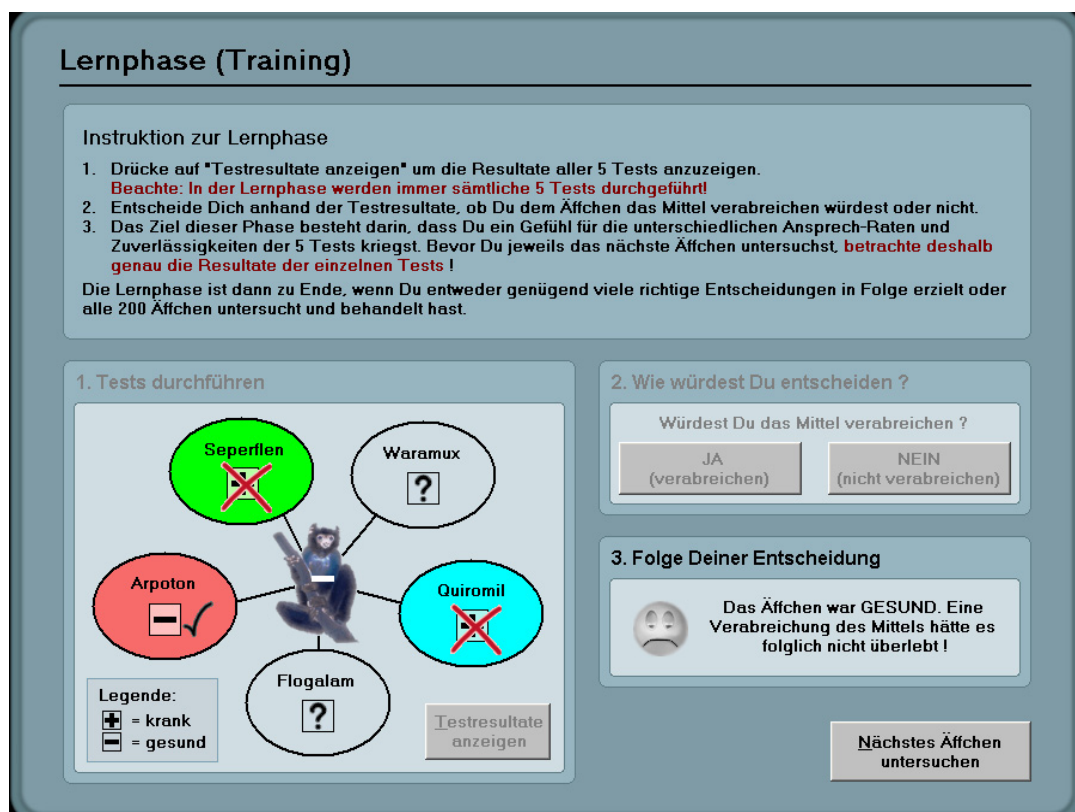


Abbildung 8.1: Screenshot der Trainingsphase. Auf der linken Seite, unterhalb der Instruktion sind die fünf Cues (Tests) zu sehen. Zur Anzeige der Testresultate muss die Schaltfläche "Testresultate anzeigen" gedrückt werden. Nach der Entscheidung für oder gegen die Verabreichung des Mittels (rechte Seite, Mitte) wird die Folge der Entscheidung mitgeteilt (rechte Seite, unten) und gleichzeitig für jeden diskriminierenden Test angezeigt, ob sein Hinweis korrekt (Häkchen) oder falsch (Kreuz) war.

8.2 Methode

8.2.1 Aufbau des Experimentes

Das Experiment umfasste eine Lern- und eine Testphase sowie eine abschliessende Nachbefragung. Für Lern- und Testphase wurde ein Information-Board Design (in Anlehnung an das *Mou-*

selab von Payne, Bettman & Johnson, 1993) verwendet (Abbildung 8.1). Im Gegensatz zur üblicherweise verwendeten tabellarischen Darstellungsform waren die Cues in diesem Experiment allerdings äquidistant auf einer gedachten Ellipse auf dem Bildschirm angeordnet. (So sollte der Eindruck einer festen Reihenfolge möglichst minimiert werden.)

Die angezeigte Information eines jeden Cues in jeder Aufgabe wurde mit der Konstruktion des Experiments einmal für alle Versuchspersonen fest generiert. Jedoch variierte die Zuordnung der Test-Bezeichnungen zu den Cues (Tabelle 8.1) von Proband zu Proband, um möglichen Salienzeffekten eines Testnamens keinen systematischen Einfluss zu erlauben. Die Validitäten und Diskriminationsraten waren innerhalb eines Blocks von 10 Aufgaben (Lernphase) bzw. 12 Aufgaben (Testphase) stets repräsentativ verteilt. Innerhalb dieser Blöcke wurde die Reihenfolge der Aufgaben zufällig bestimmt.

Tabelle 8.1: Die fünf Cues und ihre Gütwerte

Cue-Nr	V	D	Success	Usefulness
1	1.0	.10	.55	.10
2	.82	.22	.57	.18
3	.82	.61	.70	.50
4	.64	.61	.59	.39
5	.50	.80	.50	.40

Die *Cue-Struktur* für dieses Experiment war folgendermassen aufgebaut (Tabelle 8.1): Cue 1 führte immer zum korrekten Ergebnis, diskriminierte aber nur in 10% der Fälle. Die Cues 2 und 3 hatten eine gute Validität, unterschieden sich aber markant in ihren Diskriminationsraten. Die Cues 3 und 4 hatten beide eine hohe Diskriminationsrate, jedoch sehr unterschiedliche Validitäten. Cue 5 schliesslich zeichnete sich durch eine extrem hohe Diskriminationsrate aus, war im Hinblick auf die Validität jedoch völlig nutzlos. Validität und Diskriminationsrate waren also in diesem Szenario gegenläufig, und Cue 3 bot als Kombination von beiden den besten Success.

Diese Cue-Struktur war in ein Setting medizinischer Tests eingebaut. Die Versuchspersonen wurden dazu in folgende Rahmengeschichte eingeführt:

Aufgabe und Ziel

Du bist Mitglied eines zoologischen Forschungsteams, das sich in demjenigen Urwaldgebiet befindet, welches die letzten überlebenden Exemplare der vom Aussterben bedrohten Tigerschwanzäffchen beherbergt. In der Nähe dieses Urwaldes befindet sich eine Zuchtstation, die etwa 200 dieser Äffchen beherbergt. Die Grösse der im Urwald frei lebenden Population umfasst gemäss der letzten Zählung noch ca. 50 Tiere.

Leider sind die Tigerschwanzäffchen von einer heimtückischen (möglicherweise genetisch bedingten) Krankheit bedroht. Diese ist äusserlich nicht zu erkennen und führt unbehandelt innert kürzester Zeit zum Tode der betroffenen Individuen. Die Erkrankungsrate der Reservat-Population beträgt ca. 50%. Die Erkrankungsrate der Wild-Population ist nicht genau bekannt.

Erst vor kurzem wurde zufälligerweise ein Mittel (Medikament) entdeckt, welches die von der Krankheit betroffenen Äffchen in jedem Fall rettet, sofern ihnen das Mittel verabreicht wird. Unglücklicherweise

sterben die Äffchen aber, wenn sie nicht krank sind, ihnen das Mittel aber trotzdem verabreicht wird. Das Mittel hat in diesem Fall also die Wirkung eines tödlichen Giftes. Zwei Extremfälle: Würdest Du keinem Äffchen das Mittel verabreichen, so würden alle erkrankten Äffchen (50%) sterben und alle gesunden Äffchen (die anderen 50%) überleben. Im umgekehrten Fall, d.h. wenn Du das Mittel sämtlichen Äffchen verabreichen würdest, würden alle gesunden Äffchen sterben (50%) und alle erkrankten Äffchen (die anderen 50%) überleben. Die Wirkungsweise des Mittels ist bisher nicht bekannt.

Bei der Entscheidung, ob Du einem Äffchen das Mittel verabreichen sollst oder nicht, bist Du glücklicherweise nicht auf raten angewiesen. Es stehen fünf verschiedene Test-Substanzen (Tests) zur Verfügung. Diese Test-Substanzen können einem Äffchen ins Blut gespritzt werden. Dort führen sie dann mit einer bestimmten Wahrscheinlichkeit (Ansprech-Wahrscheinlichkeit) zu einer für den jeweiligen Test charakteristischen Blutbildveränderung. Jeder Test hat seine eigene (konstante) Ansprech-Wahrscheinlichkeit. Die verschiedenen Tests sprechen unterschiedlich gut an, d.h. sie funktionieren unterschiedlich gut. Wenn ein Test tatsächlich anspricht, d.h. wenn er zu einer positiven oder negativen Blutbildveränderung führt, liefert er damit einen Hinweis darauf, ob das untersuchte Äffchen krank ist oder nicht. Diese Hinweise (positive / negative Veränderungen) sind allerdings nicht in jedem Fall korrekt: Jeder der fünf Tests hat seine eigene (konstante) Zuverlässigkeit. Die fünf Tests sind voneinander unabhängig.

Zusammenfassung

- Wenn ein Test nicht anspricht, bleibt das Blutbild gleich wie zuvor (d.h. es kommt zu keiner Blutbildveränderung).
- Wenn ein Test anspricht, bedeutet dies, dass sich das Blutbild entweder positiv oder negativ verändert.
 - ➔ Eine negative Veränderung (-) deutet darauf hin, dass das Äffchen gesund ist!
 - ➔ Eine positive Veränderung (+) deutet darauf hin, dass das Äffchen krank ist!
- Die 5 Tests haben unterschiedliche Ansprech-Wahrscheinlichkeiten und Zuverlässigkeiten.

Deine Aufgabe besteht darin, möglichst viele Äffchen zu retten! Jeweils unmittelbar nach der Untersuchung eines Äffchens (mittels eines oder mehrerer Tests) musst Du Dich für oder gegen die Verabreichung des Mittels entscheiden.

Ablauf

Das Experiment besteht aus 2 Phasen: Einer Lernphase und einer Exkursionsphase:

In der Lernphase wohnst du der Untersuchung der Äffchen aus der Reservat-Population bei. Dabei sollst Du ein Gefühl für die unterschiedlichen Ansprech-Raten und Zuverlässigkeiten der fünf zur Verfügung stehenden Tests entwickeln. Dazu begleitest Du den Veterinärmediziner des Forschungsteams bei der Untersuchung der Äffchen und fällst – basierend auf den Untersuchungsergebnissen – für jedes untersuchte Äffchen eine hypothetische Behandlungsentscheid (d.h. Verabreichung vs. Nicht-Verabreichung des Mittels). Unmittelbar nach jedem Behandlungsentscheid erfährst Du, ob das Äffchen die von Dir vorgeschlagene Behandlung (Mittel vs. kein Mittel) überlebt hätte oder nicht.

In der Exkursionsphase wird es dann ernst. Dort wird Deine Aufgabe darin bestehen, möglichst viele Äffchen der Wild-Population richtig zu diagnostizieren und auch tatsächlich zu behandeln (Verabreichung vs. keine Verabreichung des Mittels), so dass möglichst viele kranke Äffchen geheilt werden und möglichst wenige gesunde Äffchen durch die Verabreichung des Mittels verenden.

In der *Lernphase* erhielten die Versuchspersonen die Möglichkeit, eine Repräsentation der über das gesamte Experiment konstant gehaltenen Diskriminationsraten und Validitäten der fünf angebotenen Cues aufzubauen. Dazu konnten sie die Arbeit von Veterinärmedizinerinnen an einer in einem grossen Freigehege gehaltenen Population beobachten. Für diese Population war bekannt, dass ca. 50% der Äffchen unter der Krankheit litten. Die Versuchsperson bekam für jedes getestete Äffchen die fünf Testresultate auf dem Bildschirm angezeigt. Sie konnte dann einen Tipp

abgeben, ob dieses Äffchen erkrankt sei oder nicht. Ihr wurde jedoch gesagt, dass dieser Tipp reine Übungszwecke habe und keinen Einfluss darauf besitze, ob das riskante Medikament appliziert werde oder nicht. Die Versuchsperson erhielt jedoch als Ergebnis rückgemeldet, ob sich das Äffchen tatsächlich als krank erwiesen hatte oder nicht. Dieses Ergebnis konnte mit dem eigenen Tipp und vor allem mit den fünf Testresultaten verglichen werden. Dann ging es zur nächsten Aufgabe.

Die Lernphase umfasste – abhängig von der Lerngeschwindigkeit der Versuchspersonen – mindestens 100, jedoch maximal 200 Durchgänge. Das Lernen wurde nach 100 korrekt getroffenen Entscheidungen abgebrochen, spätestens jedoch nach 200 Durchgängen.

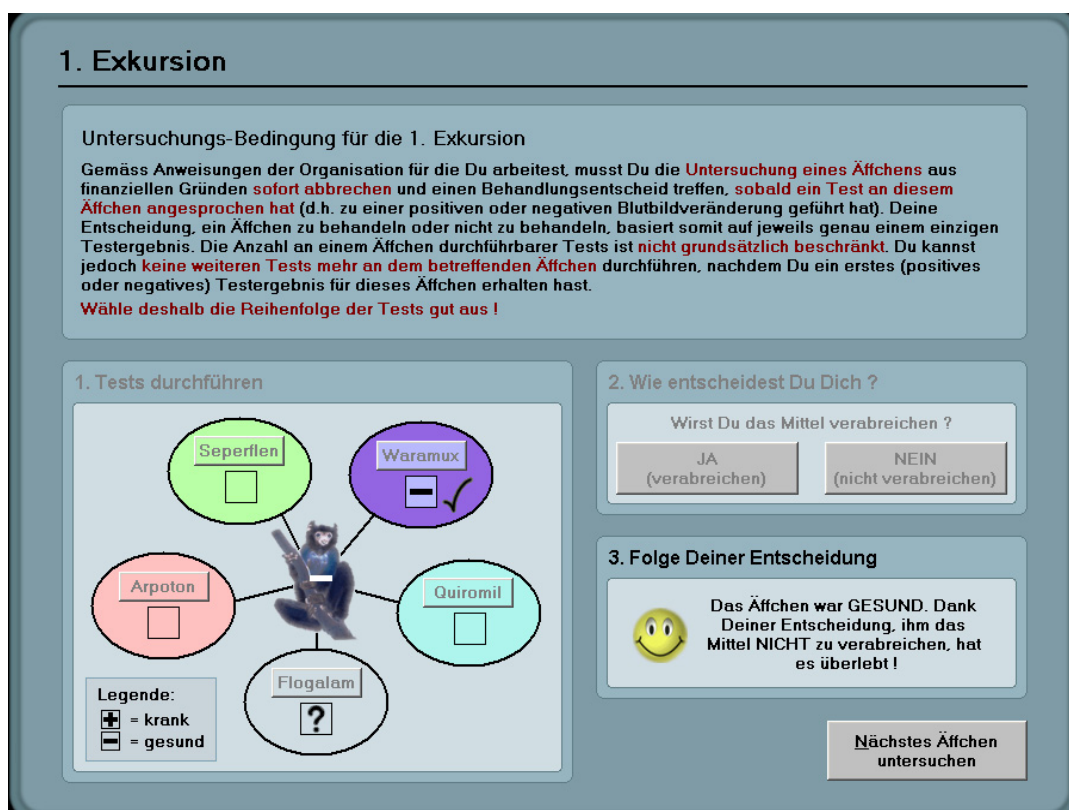


Abbildung 8.2: Screenshot der Computeroberfläche in Exkursion 1 (stellvertretend für alle drei Exkursionen). Testresultate (linke Seite, unten) werden durch Drücken der entsprechenden Schaltflächen angezeigt. Das Feedback (Häkchen oder Kreuz) für diskriminierende durchgeführte Tests wird jeweils erst nach der Entscheidung für oder gegen die Verabreichung des Mittels gegeben.

Die *Testphase* bestand aus drei Expeditionen, in denen die Versuchspersonen nun einen möglichst grossen Teil der Wildpopulation retten sollten (vgl. Abbildung 8.2). Sie konnten dazu ihre Erfahrung mit den fünf Tests einsetzen, um bei jeweils 12 eingefangenen Äffchen zu entscheiden, ob sie mit dem Medikament zu behandeln seien oder nicht. Dazu wählten die Versuchspersonen – anders als in der Lernphase – nun selbst sequentiell Tests aus. (Ihnen stand dazu pro Test eine Schaltfläche zur Verfügung, deren Anklicken das Testergebnis lieferte.) Jedoch waren die Versuchspersonen in jeder Expedition an Bedingungen hinsichtlich des Testeinsatzes gebun-

den (die bei korrekter Handhabung zur Anwendung der Validitätsordnung, des Success und der Diskriminationsrate führten). Der Abbruch der sequentiellen Testserie wurde jeweils durch die spezifische Bedingung gesteuert. Wenn die Versuchsperson keine Tests mehr zur Verfügung hatte, musste sie entscheiden, ob dem Äffchen das Medikament zu applizieren sei oder nicht.⁴¹

Die Instruktionen für die drei Expeditionen (betreffend die Bedingungen der Testanwendung) sind nachfolgend im Wortlaut wiedergegeben:

Instruktion für Expedition 1 (Anforderung Validität):

Gemäss Anweisungen der Organisation für die Du arbeitest, musst Du die Untersuchung eines Äffchens aus finanziellen Gründen sofort abbrechen und einen Behandlungsentscheid treffen, sobald ein Test an diesem Äffchen angesprochen hat (d.h. zu einer positiven oder negativen Blutbildveränderung geführt hat). Deine Entscheidung, ein Äffchen zu behandeln oder nicht zu behandeln, basiert somit auf jeweils genau einem einzigen Testergebnis. Die Anzahl an einem Äffchen durchführbarer Tests ist nicht grundsätzlich beschränkt. Du kannst jedoch keine weiteren Tests mehr an dem betreffenden Äffchen durchführen, nachdem Du ein erstes (positives oder negatives) Testergebnis für dieses Äffchen erhalten hast.

Wähle deshalb die Reihenfolge der Tests gut aus!

Instruktion für Expedition 2 (Anforderung Success):

Aus finanziellen Gründen verschärft Dein Arbeitgeber für die 2. Expedition die Untersuchungs-Bedingung: Ab sofort ist nur noch genau 1 Test pro Äffchen erlaubt. Unabhängig davon, ob der Test anspricht oder nicht, muss und kann nur genau 1 Test durchgeführt werden. Das Ziel besteht aber nach wie vor darin, möglichst viele Äffchen richtig zu behandeln (d.h. zu retten bzw. nicht umzubringen). Falls also der von Dir gewählte Test nicht ansprechen sollte, kannst Du nur noch raten, ob das Äffchen krank ist oder nicht. Deine Entscheidung, ihm das Mittel zu verabreichen oder nicht, würde damit vom Zufall abhängen. Wähle also diesen einen Test sorgfältig aus!

Instruktion für Expedition 3 (Anforderung Diskrimination):

Für die 3. Expedition lockert Dein Arbeitgeber die Untersuchungs-Bedingung nun wieder ein wenig. Bezüglich der Anzahl erlaubter Tests pro Äffchen gilt nun folgendes: Sofern der erste an einem Äffchen durchgeführte Test anspricht, darfst Du an diesem Äffchen beliebig viele weitere Tests durchführen (da nur in diesem Fall zusätzliche, rechtlich gebundene Spendengelder für die weiteren Tests zur Verfügung stehen). Falls jedoch der erste von Dir gewählte Test nicht anspricht, darfst Du keine weiteren Tests mehr durchführen. Das Überleben des betreffenden Äffchens würde damit von einer Zufallsentscheidung abhängen.

Wähle deshalb auch hier die Reihenfolge der Tests (insbesondere des ersten) sorgfältig aus!

Die Reihenfolge der in den drei Expeditionen von einer Versuchsperson gewählten Tests erlaubte folgende Auswertung: In Expedition 1 liess sich die Suchreihenfolge bis zum ersten diskriminierenden Cue betrachten. Aus Expedition 2 lag nur ein einziger Cue vor. Wenn in Expedition 3 der erste Test diskriminierte, liegen vier weitere Tests vor (deren Reihenfolge dann allerdings willkürlich ist). Bei nicht diskriminierendem erstem Test musste die Versuchsperson in dieser Exkursion raten, ob sie das Äffchen behandeln sollte oder nicht. Damit liess sich in den drei Ex-

⁴¹ Gehörte die Versuchsperson zum Drittel derjenigen Probanden, die über die Summe der drei Expeditionen die meisten Äffchen gerettet hatten, so nahm sie an einer Verlosung von 3 x 100 Schweizer Franken teil. Den Versuchspersonen wurde dieses nach der Instruktion mitgeteilt, so dass eine moderate Leistungsmotivation induziert war.

kursionen für jede Versuchsperson die Validitätsreihenfolge, der Cue mit dem höchsten Success sowie der Cue mit der höchsten Diskriminationsrate sicher bestimmen.

An die Testphase schloss sich eine *Befragung* an. Sie umfasste 3 Aufgabenpaare zur Bewertung der Cue-Güte. In Aufgabe 1a mussten die Versuchspersonen eine Rangreihe der fünf Tests (Cues) gemäss der wahrgenommenen Zuverlässigkeit (Validität) bilden. In Aufgabe 1b schätzten sie dann mittels Schieberegler für jeden der fünf Tests die absolute Ausprägung der Zuverlässigkeit prozentgenau ein. Dasselbe geschah anschliessend auch für die Usefulness⁴² (Aufgabenpaar 2a/2b) und die Diskriminationsrate (Aufgabenpaar 3a/3b). Jedes Aufgabenpaar korrespondierte so mit einer der drei Expeditionen, wobei die Aufgabenstellungen jeweils eigenständig, d.h. ohne direkten Bezug zur entsprechenden Situation (bzw. Expedition) der Anwendungsphase formuliert waren:

1. Relative Häufigkeit korrekter Testergebnisse?

- Angenommen, Du würdest jeden einzelnen Test so oft hintereinander anwenden, bis dass er insgesamt 100-mal angesprochen hat, welcher der 5 Tests würde dann am häufigsten, zweit-häufigsten, etc. ein zuverlässiges Ergebnis liefern?
- Angenommen, Du würdest den jeweils nebenstehenden Test (links) so oft anwenden, bis dass er insgesamt 100-mal angesprochen hat, wie oft (von diesen 100-mal) würde er Dir dann ein zuverlässiges Ergebnis liefern?

2. Absolute Häufigkeit korrekter Testergebnisse?

- Angenommen, Du würdest jeden einzelnen Test genau 100-mal hintereinander anwenden, welcher der 5 Tests würde dann am häufigsten, zweit-häufigsten, etc. ein zuverlässiges Ergebnis liefern?
- Angenommen, Du würdest den jeweils nebenstehenden Test (links) genau 100-mal hintereinander anwenden, wie oft (d.h. in wie vielen dieser 100 Testanwendungen) würde er Dir dann ein zuverlässiges Ergebnis liefern?

3. Ansprech-Häufigkeit der Tests?

- Angenommen, Du würdest jeden einzelnen Test genau 100-mal hintereinander anwenden, welcher der 5 Tests würde dann am häufigsten, zweit-häufigsten, etc. ansprechen?
- Angenommen, Du würdest den jeweils nebenstehenden Test (links) genau 100-mal hintereinander anwenden, wie oft (d.h. in wievielen dieser 100 Testanwendungen) würde er dann überhaupt ansprechen (d.h. eine korrekte oder falsche Vorhersage machen)?

⁴² Wie bereits in der Einleitung erwähnt, sind Verhaltensdaten und Befragungsdaten hinsichtlich Success (Verhalten) und Usefulness (Befragung) dissoziiert. Dies erlaubt Aussagen über beide integrierte Gütemasse und somit eine Erweiterung der referierten Erkenntnisse von Newell et al. (2003). Mit der Befragung sollen also nicht in erster Linie die Resultate der Anwendungsphase überprüft werden, sondern das Ziel besteht in einer breiteren Datenbasis für das bei den Vpn vorhandene Verständnis der Cuegüte. Aufgrund der Instruktion für Expedition 2 (die Success als Norm implizierte) umfasste die aus dieser Expedition resultierende Datenspur jeweils lediglich einen Cue pro Durchgang. Wichtig war uns deshalb in der Befragung, dass die Vpn nicht nur einen einzigen Cue als integriertes Gütemass auswählen, sondern sämtliche 5 Cues bezüglich einer Suchstrategie rangieren, die V und D kombiniert.

8.3 Geplante Auswertung

Für die Auswertung ermöglichen sich damit sechs verschiedene Kriterien für das Gelingen des Cue-Lernens:

1. *Die Cue-Güte des zuerst gewählten Tests:* Zur Bestimmung der Abweichung des Verhaltens der Versuchsperson von rein zufälligem Verhalten ist die Null-Hypothese "Die Vp hat nichts gelernt" mittels einseitigem One-Sample-t-Test zu überprüfen: Für jede Expedition (V, Suc, D) wird der expeditionsabhängige mittlere Gütewert der fünf Tests (Validität [V]: .756; Success [Suc]: .580; Diskriminationsrate [D]: .468) mit den mittleren Gütewerten der von der Versuchsperson zuerst gewählten Tests verglichen.
2. *Der "Erfüllungs-Score" (ES) pro Expedition und Versuchsperson:* Der Gütewert (V, Suc, D) des zuerst durchgeführten Tests wird dabei in jeder Aufgabe in Beziehung gesetzt zum Gütewert des normativ in dieser Expedition anzuwendenden Tests. Für eine einzelne Aufgabe gilt Erfüllungswert = 1, wenn eine Übereinstimmung mit der normativen Testanwendung besteht. Wird der normativ schlechteste Test gewählt, dann gilt Erfüllungswert = 0. Zwischen diesen beiden Extremen wird ein Messwert linear zur Skala des normativ gerade geltenden Gütewertes (V, Suc, D) berechnet. Der mittlere Erfüllungswert über die 12 Aufgaben ergibt für diese Person in dieser Bedingung den Erfüllungs-Score.
3. *Adaptivität des gezeigten Cue-Suchverhaltens auf der Basis der tatsächlichen Cue-Güte:* Es wäre denkbar, dass eine Versuchsperson in jeder der drei Expeditionen dasselbe Verhaltensmuster zeigt, dass sie also ihr Verhalten nicht der entsprechenden Situation anpasst, sondern situationsunabhängig immer dieselben Cues bevorzugt. Deshalb berechnen wir für jede Expedition zusätzlich auch die beiden Erfüllungs-Scores, die sich ergeben, wenn die beiden jeweils gerade nicht geforderten Normen angewendet werden. Wenn ein Proband z.B. in jeder Expedition die validesten Cues bevorzugen würde, wären für diese Person die sich durch den Transfer der Verhaltensnorm von Expedition 1 auf das Verhalten in Expedition 2 und Expedition 3 ergebenden Erfüllungsscores grösser als die originalen Erfüllungsscores aus den Expeditionen 2 und 3.
4. Erfüllungs-Scores sowie Adaptivität des gezeigten Cue-Suchverhaltens auf der Basis der subjektiven Überzeugungen bezüglich der Validitäten und Diskriminationsraten der Tests (analog Auswertung 2 und 3). Diese Auswertung trägt jedoch nur dann zum Verständnis des Verhaltens der Versuchspersonen im Experiment bei, wenn die Erfüllungs-Scores hinsichtlich der im Befragungsteil angegebenen Werte besser ausfallen als mit den tatsächlichen Werten berechnet.
5. Zur Beurteilung des Lernens auf deklarativer Ebene lassen sich die von den Versuchspersonen im Befragungsteil angegebenen Rangordnungen und Ausprägungswerte mit den tatsächlichen Rängen der Tests (ordinal) und mit der tatsächlichen Cuegüte (Pearson-Korrelationen) vergleichen. Daneben ist für die angegebene Cue-Güte auch eine Abweichung zu den tatsächlichen Werten als Differenzwert zu berechnen.

6. Für jede Versuchsperson wird die Anzahl korrekter Entscheidungen pro Expedition miteinander verglichen. Dabei ist zu berücksichtigen, dass die maximale Anzahl der erreichbaren korrekten Entscheidungen (bei konsequenter Anwendung der jeweils normativ korrekten Suchstrategie) nicht bei allen 3 Expeditionen gleich gross war: In Expedition 1 hätte die konsequente Anwendung der V-Strategie zu 12 korrekten und 0 falschen Entscheidungen geführt. In Expedition 2 hätte der Kauf des jeweils ersten Cues pro Aufgabe (bei konsequenter Anwendung der Success-Strategie) 6-mal zu einem korrekten Tipp geführt und 1-mal zu einem falschen. Zudem hätte man in fünf Fällen keinen Tipp erhalten und raten müssen (Ratewahrscheinlichkeit = 50%). Im Mittel würde man somit lediglich 8 bis 9 korrekte Entscheidungen treffen. In Expedition 3 wiederum würde ein Proband bei konsequenter Anwendung der D-Strategie 10 Mal einen korrekten Tipp erhalten, da der am stärksten diskriminierende Test ($D = .80$) in 10 der 12 Aufgaben anschlägt. In diesen 10 Fällen könnte die Versuchsperson auch die restlichen 4 Cues zu Rate ziehen und bei normativ korrektem Vorgehen (d.h. ihre Entscheidung auf dem validesten Test basierend) 10 Mal eine korrekte Entscheidung treffen. In den verbleibenden 2 Fällen müsste sie raten (Ratewahrscheinlichkeit = 50%) und käme damit im Mittel auf 11 korrekte Entscheidungen.

8.3.1 Stichprobe

25 Versuchspersonen aus einer bezüglich Alter, Bildung und Geschlecht (11 Männer und 14 Frauen) gut durchmischten Stichprobe nahmen am Experiment teil. Das Durchschnittsalter lag bei 34.6 Jahren (Streubreite 23 – 62 Jahre, $SD = 10.4$ Jahre). Eine Versuchsperson (Vp 11) musste von der Datenauswertung ausgeschlossen werden, da sie (laut eigenen Angaben in der Nachbefragung) einen Teil der Instruktionen völlig falsch verstanden hatte.

8.4 Resultate

8.4.1 Auswertung 1 und 2: Lernen des besten Cues (One-Sample-T-Test, einseitig) und Lernfähigkeit – Erfüllungsscores bei tatsächlichem Verhalten

Die Auswertungen 1 und 2 (Tabellen 8.2 bis 8.4) beziehen sich auf den jeweils ersten eingesetzten Test. Sie korrespondieren sehr stark und werden deswegen gemeinsam dargestellt. Während der One-sample t-Test ein individuelles inferenzstatistisches Mass darstellt (wir wählen als Ablehnungskriterium für die Null-Hypothese „nicht gelernt“ das 1%-Niveau), haben die Erfüllungsscores auf individueller Ebene deskriptiven Charakter (sie sollten $> .70$ ausfallen, damit von „gelernt“ gesprochen werden kann, da die Signifikanzgrenze im T-Test praktisch perfekt mit einem Erfüllungsscore von $.70$ zusammenfällt).

18 der 24 Versuchspersonen sind mit der Lernaufgabe gut bis perfekt klargekommen und weisen durchgängig Erfüllungsscores von zumindest $> .50$ auf. Die Vpn 3, 5, 12, 14, 18 hingegen scheinen die Cue-Güte nicht oder nur sehr partiell gelernt zu haben: Die Vpn 3 und 12 haben lediglich gelernt, dass ein Cue immer diskriminiert, und nehmen in der Testphase unabhängig vom Setting mehrheitlich diesen. Vp 14 zeigt ein inkonsistentes, nicht aufgabenbezogenes Muster. Vp 18 zeigt ein etwas aufgabenbezogeneres Muster, hat jedoch den jeweils besten Cue nicht gelernt. Vp

5 schliesslich zeigt ein sehr aufgabenbezogenes Muster, wendet den jeweils besten Cue jedoch nicht konsequent genug an (ist sich vermutlich unsicher).

Tabelle 8.2: Expedition 1 (Validität) – One-Sample-T-Test und Erfüllungs-Scores (ES). Die letzte Spalte gibt die Anzahl vollständig validitätskonform gelöster Aufgaben an (max. = 12). P-Werte unter 1% sind fett gedruckt. ES: Werte > .70 sind fett gedruckt.

VpID	Validität des jeweils ersten pro Aufgabe durchgeführten Tests												Validität (Testwert=75.6)			Voll V-konf.
	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	Mean	p	ES (%)	
1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.0000	1.00	12
2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.0000	1.00	9
4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.0000	1.00	12
8	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.0000	1.00	12
10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.0000	1.00	9
16	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.0000	1.00	5
17	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.0000	1.00	1
21	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.0000	1.00	6
23	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.0000	1.00	8
6	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.82	1.00	1.00	1.00	1.00	.99	.0000	.97	11
9	1.00	1.00	.82	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.99	.0000	.97	11
22	.82	.82	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.97	.0000	.94	6
7	.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.96	.0003	.92	11
25	1.00	.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.96	.0003	.92	11
15	.82	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.82	.82	.82	1.00	.94	.0000	.88	5
19	1.00	.82	.82	.82	.82	.82	.82	1.00	.82	1.00	1.00	1.00	.90	.0001	.79	2
24	.82	.82	.82	.82	.82	.82	.82	1.00	1.00	1.00	1.00	1.00	.90	.0001	.79	5
20	.82	1.00	.82	.82	1.00	.82	.82	.82	1.00	.82	.82	.82	.87	.0004	.73	2
14	.82	.50	.82	.64	1.00	.64	1.00	1.00	1.00	.64	1.00	1.00	.84	.0790	.68	1
5	.64	.82	.82	.82	1.00	.82	1.00	.64	1.00	.82	.82	.82	.84	.0220	.67	0
13	.82	.82	.82	.82	.82	.82	.82	.82	.82	.82	.82	.82	.82	.0000	.64	0
18	.82	.82	.64	.82	.64	.82	.82	.82	.50	.82	.82	.82	.76	.4093	.53	0
3	.50	.64	.82	.50	.82	.82	.50	.64	.82	.82	.50	.50	.66	> .50	.31	0
12	1.00	.50	.50	.50	.50	.50	.50	.50	.50	.50	.50	.50	.54	> .50	.08	0

In der Validitätsbedingung startet Vp 13 konsequent mit dem zweitvalidesten Cue, vier weitere Vpn starten mal mit dem validesten, mal mit dem zweitvalidesten Cue (sind sich des validesten also nicht sicher). Gemäss dem Kriterium des 1%-Niveaus beim t-Test haben diese fünf Personen Validität gelernt. Alle anderen 14 Vpn haben den validesten Cue erkannt und durchgängig (oder mit ein oder zwei Ausnahmen) als ersten Test angewendet.

In der Success-Bedingung schwanken Vp 9 und Vp 16 zwischen den beiden besten Cues, Vp 17 und Vp 19 wenden den besten Cue nur fünf mal an, Vp 20, Vp 21 und Vp 25 nur vier mal. Bei Vp 23 (durchgängig den meistdiskriminierenden Cue) ist unklar, ob sie die Testbedingung richtig verstanden hat (sowohl Validität als auch Diskriminationsrate bearbeitet sie mit einem Erfüllungsscore von 1.00). Bei mindestens 11 Versuchspersonen darf man also bereits davon ausgehen, dass sie den Success der Cues richtig repräsentiert haben. Ferner sind die Testanwendungen der Vpn 9 und Vp 16 im t-Test auf dem 1%-Niveau signifikant, also ebenfalls hinzuzuzählen.

Tabelle 8.3: Expedition 2 (Success) – One-Sample-T-Test und Erfüllungs-Scores (ES). P-Werte unter 1% sind fett gedruckt. ES: Werte > .70 sind fett gedruckt.

VpID	Success des jeweils ersten pro Aufgabe durchgeführten Tests												Success (Testwert=58.02)		
	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	Mean	p	ES (%)
2	.70	.70	.70	.70	.70	.70	.70	.70	.70	.70	.70	.70	.70	.0000	1.00
4	.70	.70	.70	.70	.70	.70	.70	.70	.70	.70	.70	.70	.70	.0000	1.00
8	.70	.70	.70	.70	.70	.70	.70	.70	.70	.70	.70	.70	.70	.0000	1.00
10	.70	.70	.70	.70	.70	.70	.70	.70	.70	.70	.70	.70	.70	.0000	1.00
13	.70	.70	.70	.70	.70	.70	.70	.70	.70	.70	.70	.70	.70	.0000	1.00
22	.70	.70	.70	.70	.70	.70	.70	.70	.70	.70	.70	.70	.70	.0000	1.00
6	.70	.70	.70	.70	.70	.70	.59	.70	.70	.70	.70	.70	.69	.0000	.95
15	.55	.70	.70	.70	.70	.70	.70	.70	.70	.70	.70	.70	.68	.0000	.94
7	.70	.70	.70	.70	.70	.70	.70	.59	.59	.70	.70	.70	.68	.0000	.91
1	.57	.70	.70	.70	.59	.70	.70	.70	.70	.70	.70	.70	.68	.0000	.90
24	.70	.70	.70	.70	.50	.50	.70	.70	.70	.70	.70	.70	.66	.0016	.83
16	.70	.59	.59	.70	.59	.70	.59	.70	.70	.70	.70	.70	.66	.0002	.81
9	.70	.70	.70	.70	.70	.59	.59	.59	.59	.70	.59	.59	.64	.0020	.72
17	.70	.70	.70	.70	.70	.70	.55	.50	.70	.55	.55	.55	.63	.0273	.67
25	.70	.70	.57	.59	.70	.59	.59	.59	.59	.70	.59	.59	.62	.0136	.62
21	.70	.70	.55	.55	.70	.70	.59	.59	.59	.59	.59	.59	.62	.0310	.59
20	.57	.55	.57	.70	.55	.57	.70	.55	.70	.57	.70	.55	.61	.1114	.54
5	.57	.55	.70	.50	.59	.55	.59	.70	.57	.70	.50	.70	.60	.2003	.51
19	.70	.50	.70	.50	.55	.70	.50	.55	.70	.50	.55	.70	.59	.3082	.48
14	.70	.55	.59	.59	.57	.50	.70	.55	.59	.50	.55	.59	.58	.4818	.41
18	.50	.50	.50	.50	.50	.57	.70	.50	.59	.50	.70	.55	.55	> .50	.25
3	.59	.59	.50	.50	.50	.50	.50	.50	.50	.50	.50	.50	.51	> .50	.07
12	.50	.50	.50	.50	.50	.50	.50	.50	.50	.50	.50	.50	.50	> .50	.00
23	.50	.50	.50	.50	.50	.50	.50	.50	.50	.50	.50	.50	.50	> .50	.00

In der Diskriminationsbedingung wählen die Vpn 1, 13, 15, 16, 17 und 24 konsequent den am zweitmeisten diskriminierenden Cue aus. Die Vpn 19, 21 und 22 schwanken zwischen den beiden meistdiskriminierenden Cues. Diese Personen haben also grundsätzlich die Diskriminationsrate gelernt, allerdings nicht unbedingt den meistdiskriminierenden Cue. Lediglich von den drei Vpn 14, 18 und 20 muss angenommen werden, dass sie die Diskriminationsrate nicht gelernt haben. Die Vpn 12 und 17 verfehlen das 1%-Signifikanzniveau, weil sie jeweils in einer einzigen Aufgabe gleich mit dem validesten Cue gestartet sind. Es ist also Geschmackssache, sie zu denjenigen zu zählen, die die Diskriminationsrate gelernt haben, oder nicht.

Damit können die Auswertungen 1 und 2 so zusammengefasst werden, dass Diskriminationsrate und Validität von der grossen Mehrzahl der Versuchspersonen (jeweils mindestens 18 der 24) gelernt wurde, Success hingegen nur etwa von der Hälfte (13 der 24). Fünf der 24 Probanden sind insgesamt schlecht mit dem Lernen und Anwenden zurechtgekommen.

Tabelle 8.4: Expedition 3 (Diskriminationsrate) – One-Sample-T-Test und Erfüllungs-Scores (ES).
P-Werte unter 1% sind fett gedruckt. ES: Werte > .70 sind fett gedruckt.

VpID	Diskriminationsrate des jeweils ersten durchgeführten Tests												Diskr.-Rate (Testw.=46.8)		
	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	Mean	p	ES (%)
2	.80	.80	.80	.80	.80	.80	.80	.80	.80	.80	.80	.80	.80	.0000	1.00
3	.80	.80	.80	.80	.80	.80	.80	.80	.80	.80	.80	.80	.80	.0000	1.00
4	.80	.80	.80	.80	.80	.80	.80	.80	.80	.80	.80	.80	.80	.0000	1.00
6	.80	.80	.80	.80	.80	.80	.80	.80	.80	.80	.80	.80	.80	.0000	1.00
7	.80	.80	.80	.80	.80	.80	.80	.80	.80	.80	.80	.80	.80	.0000	1.00
9	.80	.80	.80	.80	.80	.80	.80	.80	.80	.80	.80	.80	.80	.0000	1.00
23	.80	.80	.80	.80	.80	.80	.80	.80	.80	.80	.80	.80	.80	.0000	1.00
25	.80	.80	.80	.80	.80	.80	.80	.80	.80	.80	.80	.80	.80	.0000	1.00
10	.61	.80	.80	.80	.80	.80	.80	.80	.80	.80	.80	.80	.78	.0000	.98
8	.80	.80	.80	.80	.80	.80	.80	.80	.22	.80	.80	.80	.75	.0001	.93
22	.80	.80	.80	.80	.80	.61	.61	.61	.61	.61	.61	.61	.69	.0000	.84
5	.61	.80	.61	.80	.22	.61	.61	.61	.80	.80	.80	.61	.66	.0012	.80
19	.61	.80	.61	.61	.80	.61	.80	.80	.22	.61	.80	.61	.66	.0012	.80
12	.80	.80	.61	.61	.80	.10	.80	.80	.10	.80	.80	.80	.65	.0184	.79
21	.61	.80	.61	.61	.61	.61	.80	.61	.61	.61	.61	.61	.64	.0000	.77
1	.61	.61	.61	.61	.61	.61	.61	.61	.61	.61	.61	.61	.61	.0000	.73
13	.61	.61	.61	.61	.61	.61	.61	.61	.61	.61	.61	.61	.61	.0000	.73
15	.61	.61	.61	.61	.61	.61	.61	.61	.61	.61	.61	.61	.61	.0000	.73
16	.61	.61	.61	.61	.61	.61	.61	.61	.61	.61	.61	.61	.61	.0000	.73
24	.61	.61	.61	.61	.61	.61	.61	.61	.61	.61	.61	.61	.61	.0000	.73
18	.80	.80	.80	.61	.22	.61	.80	.61	.61	.10	.22	.80	.58	.0780	.69
17	.61	.61	.61	.61	.61	.61	.61	.10	.61	.61	.61	.61	.57	.0195	.67
14	.80	.61	.10	.10	.10	.61	.61	.10	.61	.10	.61	.61	.41	> .50	.45
20	.22	.22	.22	.61	.61	.10	.10	.22	.10	.61	.80	.22	.34	> .50	.34

8.4.2 Auswertung 3: Adaptivität des Suchverhaltens: Erfüllungsscores bei transferiertem Verhalten

Nachdem Auswertung 1 und 2 gezeigt haben, dass die Mehrzahl der Personen hinsichtlich Validität, Diskriminationsrate und Success gelernt haben, werden in dieser Auswertung die Erfüllungs-Scores des individuell in den drei Testphasen gezeigten Verhaltens auf die beiden jeweils anderen Bedingungen projiziert. Adaptives Verhalten liegt dann vor, wenn der Erfüllungsgrad für eine Versuchsperson jeweils der geforderten Bedingung besser entspricht als den beiden anderen Bedingungen. Bei Verletzungen der Adaptivität kann etwas darüber ausgesagt werden, welche(s) der Cue-Gütemasse die betreffende Versuchsperson vornehmlich einsetzt.

Tabelle 8.5 zeigt, dass das validitätsorientierte Verhalten auf die Validitäts-Bedingung begrenzt bleibt und nicht auf die beiden anderen Exkursionen übertragen wird. Die Diskriminationsrate wird jedoch häufig in der Success-Bedingung eingesetzt (von 10 der 24 Probanden), und 7 Probanden übertragen umgekehrt die Success-Folge auf die Diskriminations-Bedingung.

Vier der 24 Probanden handeln in den drei Exkursionen nicht adaptiv: Vp 3 und Vp 12 wenden immer die Diskriminationsrate an, Vp 13 die Success-Folge und Vp 20 die Validität. Von den anderen Versuchspersonen darf man annehmen, dass sie, sofern sie gelernt haben, mehr als nur

ein Gütemass aufgenommen und eingesetzt haben. Sieben Probanden verhalten sich hingegen perfekt adaptiv und wenden die jeweils geforderte Strategie in den richtigen Bedingungen an.

Tabelle 8.5: Adaptivität der Versuchspersonen – tatsächliche und theoretische Erfüllungsscores bei transferiertem Verhalten. Der jeweils höchste Erfüllungsscore pro Bedingung und Versuchsperson ist fett gedruckt.

VpID	Validität			Success			Diskriminationsrate			Adaptivität
	V	S_to_V	D_to_V	V_to_S	S	D_to_S	V_to_D	S_to_D	D	
1	1.00	.26	.00	.61	.90	.68	.64	1.00	.73	
2	1.00	.26	.00	.64	1.00	.73	.00	.00	1.00	adaptiv
3	.31	.49	.84	.05	.07	.95	.00	.00	1.00	nur D
4	1.00	.26	.00	.64	1.00	.73	.00	.00	1.00	adaptiv
5	.67	.56	.41	.53	.51	.56	.34	.48	.80	
6	.97	.26	.01	.61	.95	.73	.00	.00	1.00	adaptiv
7	.92	.23	.08	.58	.91	.73	.00	.00	1.00	adaptiv
8	1.00	.26	.00	.64	1.00	.73	.05	.03	.93	adaptiv
9	.97	.26	.01	.46	.72	.73	.00	.00	1.00	
10	1.00	.26	.00	.64	1.00	.73	.02	.04	.98	adaptiv
12	.08	.02	.92	.00	.00	1.00	.27	.21	.79	nur D
13	.64	1.00	.73	.64	1.00	.73	.64	1.00	.73	nur S
14	.68	.40	.39	.50	.41	.55	.68	.51	.45	
15	.88	.45	.20	.67	.94	.67	.64	1.00	.73	
16	1.00	.26	.00	.52	.81	.73	.58	.91	.73	
17	1.00	.26	.00	.71	.67	.51	.67	.94	.67	
18	.53	.61	.57	.27	.25	.78	.31	.27	.69	
19	.79	.69	.43	.52	.48	.64	.37	.53	.80	
20	.73	.44	.22	.76	.54	.30	.65	.42	.34	nur V
21	1.00	.26	.00	.52	.59	.61	.53	.83	.77	
22	.94	.38	.12	.64	1.00	.73	.37	.58	.84	adaptiv
23	1.00	.26	.00	.00	.00	1.00	.00	.00	1.00	
24	.79	.48	.24	.53	.83	.77	.64	1.00	.73	
25	.92	.23	.08	.43	.62	.68	.00	.00	1.00	

Die Vpn 1, 15, 16 und 24 verwenden die Success-Strategie, wenn die Diskriminationsrate gefordert wäre, bei den Vpn 5, 9, 18, 19, 23 und 25 ist es genau andersherum.

Nur zwei der Probanden zeigen ein sehr idiosynkratisches Verhalten: Vp 17 wendet die Success-Strategie dort an, wo Diskriminationsrate gefordert wäre. In der Success-Bedingung selbst hingegen entspricht ihr Verhalten am ehesten noch der Validitätsfolge. Vp 14 erreicht grundsätzlich nur niedrige Erfüllungsscores.

Was sie an Validität gelernt hat (und auch in der entsprechenden Bedingung einsetzt), überträgt sie exakt auf die Exkursion 3. Ihr Verhalten in der Success-Bedingung lässt sich mit keiner der drei Strategien gut beschreiben.

8.4.3 Auswertung 4: Erfüllungsscores und Adaptivität des Suchverhaltens basierend auf der subjektiven Cue-Güte

Es ist ja anzunehmen, dass die Probanden die Testaufgaben hinsichtlich ihrer gelernten Repräsentation der Cue-Qualitäten bearbeitet haben. Falls nun die nach dem Experiment erhobenen subjektiven Angaben über Validitäten und Diskriminationsraten zu signifikant höheren Erfüllungsscores führen als die Berechnungen auf der Ebene der tatsächlichen Gütewerte der Cues, kann die Wiederholung der Auswertungen 2 und 3 zum Verständnis des Verhaltens der Versuchspersonen beitragen. Zwar steigen die Erfüllungsscores auf der Basis der Daten der Nachbefragung tatsächlich im Mittel ein wenig an (für Validität $ES = .88$ gegenüber $ES = .83$, $p = .63$, für Diskriminationsrate $ES = .83$ gegenüber $ES = .82$, $p = .347$), jedoch sind diese Unterschiede im T-Test für abhängige Stichproben nicht signifikant. Da eine Durchsicht der Berechnungen für keine Versuchsperson zusätzliche Erkenntnisse bot, wird auf eine detaillierte Darstellung von Resultaten auf individuellem Niveau an dieser Stelle verzichtet.

Tabelle 8.6: Korrelativer Zusammenhang der expliziten Schätzungen der Versuchspersonen mit den tatsächlichen Gütewerten, aufgeschlüsselt nach Expeditionen.

VpID	Pearson			Spearman		
	V	V*D	D	V	V*D	D
1	.87	.84	.94	.975	.900	.975
2	.87	-.90	.95	.975	.100	.975
3	-.93	.99	.99	-.925	.900	.975
4	.95	.80	.99	.975	.900	.975
5	.91	.36	-.92	.825	.900	-.825
6	.95	.71	.97	.975	.900	.975
7	.97	.71	.98	.975	.700	.975
8	.95	.16	.99	.975	.300	.975
9	.93	.91	.96	.975	.800	.975
10	.73	.70	.96	.725	.800	.975
12	.37	.94	.97	.475	.900	.975
13	-.42	.78	.99	-.275	.600	.975
14	-.24	.60	.08	-.275	.700	.125
15	.36	-.13	-.39	.475	-.100	-.625
16	.69	.28	.75	.675	.300	.825
17	.41	.60	.56	.625	.700	.675
18	.13	.37	.96	.175	-.300	.975
19	.52	-.01	.96	.625	.000	.825
20	.94	-.71	-.87	.725	-.600	-.825
21	.42	.93	.92	.425	.900	.825
22	.98	.50	.94	.975	.400	.975
23	.86	.85	.95	.875	.900	.975
24	.94	.79	.96	.975	.700	.975
25	.99	.93	.96	.975	.900	.975

Tabelle 8.7: Mittlere Differenzen zwischen den Schätzungen der Cuegüte (Nachbefragung) und den tatsächlichen Cuegüten. Bei den vorzeichen-behafteten Differenzen bezeichnen negative Werte eine systematische Unterschätzung, positive Werte eine Überschätzung der tatsächlichen Cuegüte.

Vp	Vorzeichenbehafteter Schätzfehler			Absoluter Schätzfehler		
	V	V*D	D	V	V*D	D
1	-.05	.15	.09	.07	.17	.11
2	-.11	.08	.06	.11	.17	.11
3	-.30	.12	-.04	.41	.12	.04
4	-.10	.32	.09	.10	.32	.09
5	-.13	.13	.00	.13	.13	.42
6	-.03	.07	.08	.07	.09	.08
7	.01	.18	.05	.04	.18	.05
8	-.03	.29	.06	.05	.29	.06
9	-.07	.23	.03	.07	.23	.06
10	-.19	.19	.00	.19	.22	.06
12	-.24	.03	.01	.27	.10	.07
13	-.32	.02	-.04	.32	.11	.07
14	-.36	.17	-.03	.36	.17	.30
15	-.15	.29	.16	.20	.31	.38
16	-.31	.18	.07	.31	.34	.15
17	-.14	-.07	-.02	.25	.10	.22
18	-.12	.02	.23	.20	.15	.23
19	-.07	.38	.09	.16	.38	.09
20	-.36	.03	-.35	.36	.38	.42
21	.16	.24	.16	.16	.24	.17
22	-.14	.20	.15	.14	.24	.16
23	.03	.11	.07	.06	.11	.07
24	-.08	.16	.14	.08	.16	.14
25	-.13	.21	.11	.13	.21	.12
Mean	-.13	.15	.05	.18	.21	.15
SD	.12	.11	.11	.11	.09	.11

8.4.4 Auswertung 5: Abschätzung des Lernens anhand deklarativer Angaben (Rangordnungen, Pearson-Korrelationen und Differenzbeträge)

Aus Tabelle 8.6 sind die Korrelationskoeffizienten für jede Expedition getrennt ersichtlich.⁴³ Zur besseren Orientierung sind dabei Werte $> .70$ hervorgehoben (in diesen Fällen liegt zumindest mehr als 50% Varianzaufklärung vor. Die Diskriminationsrate ist von 18 der 24 Vpn mit einer Pearson-Korrelation von $> .70$ gelernt worden, bei der Validität sind es 13 Personen und bei Usefulness acht. Die Diskriminationsrate wurde also sehr gut gelernt, bei Usefulness hingegen hatte die Mehrzahl der Probanden Mühe. Bei den Rangkorrelationen ist der Befund ähnlich – lediglich

⁴³ Die maximal mögliche Rangkorrelation zwischen den eingeschätzten und den tatsächlichen Cue-Rängen beträgt 1.00. Aufgrund der gebundenen Rangplätze bei Validität und Diskriminationsrate (vgl. Tabelle 1) ergibt sich in den betreffenden Spalten jedoch rein rechnerisch lediglich ein Wert von .975. Dieser Wert entspricht somit der maximalen Korrelation.

an den Grenzen des willkürlich festgesetzten Kriteriums von .70 kommt es bei der einen oder anderen Person zu Verschiebungen.

Die Korrelationswerte geben nicht den Blick auf die systematischen Abweichungen der Einschätzungen der Cue-Güte von den tatsächlichen Werten frei (Tabelle 8.7). Auch wenn für non-kompensatorisches Verhalten (etwa im Sinne von Take-the-Best) lediglich ein Wissen um die Rangordnung von Bedeutung ist, sollte doch festgehalten werden, dass die Validitäten der Cues systematisch unterschätzt werden (von 21 der 24 Versuchspersonen), während hingegen die Diskriminationsraten (von 18 Probanden) und die Usefulness-Raten von (23 Probanden) mehrheitlich überschätzt werden. Im Durchschnitt liegen die Schätzungen bei den Diskriminationsraten um 15% daneben, bei den Validitäten um 17% und bei der Usefulness sogar um 21%.

Abbildung 8.8: Anzahl korrekter Entscheidungen (Treffer) pro Versuchsperson und Expedition; zusätzlich zu den absoluten Trefferzahlen sind in Klammern die Quoten im Vergleich mit der Erfolgserwartung bei instruktionsgemäßem Verhalten angegeben. Diejenigen Bedingungen, die hinsichtlich der Erfüllungs-Scores adaptiv durchgeführt wurden (Auswertung 3), sind fett gedruckt.

VpID	Exp1: V	Exp2: Suc	Exp3: D	Summe
4	12 (100%)	9 (106%)	11 (100%)	32 (306%)
6	12 (100%)	8 (94%)	11 (100%)	31 (294%)
9	12 (100%)	8 (94%)	11 (100%)	31 (294%)
7	11 (92%)	8 (94%)	11 (100%)	30 (286%)
13	10 (83%)	10 (118%)	9 (82%)	29 (283%)
10	11 (92%)	8 (94%)	10 (91%)	29 (277%)
1	12 (100%)	8 (94%)	9 (82%)	29 (276%)
15	10 (83%)	9 (106%)	9 (82%)	28 (271%)
8	11 (92%)	8 (94%)	9 (82%)	28 (268%)
24	11 (92%)	8 (94%)	9 (82%)	28 (268%)
25	11 (92%)	7 (82%)	10 (91%)	28 (265%)
22	11 (92%)	8 (94%)	8 (73%)	27 (259%)
19	9 (75%)	9 (106%)	8 (73%)	26 (254%)
23	11 (92%)	5 (59%)	11 (100%)	27 (250%)
16	9 (75%)	8 (94%)	8 (73%)	25 (242%)
2	7 (58%)	8 (94%)	9 (82%)	24 (234%)
17	8 (67%)	8 (94%)	8 (73%)	24 (234%)
3	9 (75%)	7 (82%)	8 (73%)	24 (230%)
21	9 (75%)	8 (94%)	6 (55%)	23 (224%)
18	9 (75%)	5 (59%)	9 (82%)	23 (216%)
5	6 (50%)	6 (71%)	9 (82%)	21 (202%)
20	9 (75%)	6 (71%)	6 (55%)	21 (200%)
12	8 (67%)	4 (47%)	9 (82%)	21 (196%)
14	6 (50%)	8 (94%)	5 (45%)	19 (190%)

8.4.5 Auswertung 6: Entscheidungsebene (Anzahl korrekter Entscheidungen)

Tabelle 8.8 zeigt sowohl die absolute als auch die normierte Anzahl der korrekten Entscheidungen pro Versuchsperson und Expedition. Die Normierung basiert auf der für die jeweilige Expedition zu erwartenden mittleren Anzahl Treffer bei normativem Verhalten (d.h. bei Anwendung von V, Suc und D und bei perfekt gelernten Cue-Rangordnungen): Zu erwarten sind 12 Treffer für Expedition 1, 8.5 Treffer für Expedition 2 und 11 Treffer für Expedition 3. Aufgrund dieser Normierungsweise sind – wenn die Vp bei den Rateentscheidungen Glück hatte – in Exp. 2 und 3 prozentuale Trefferraten $> 100\%$ möglich.

Die Absolutwerte fallen in der Validitätsbedingung höher aus (durchschnittlich 9.7 Treffer gegenüber 8.9 bei der Diskriminationsrate und 7.5 in der Success-Bedingung), jedoch ist dies mit den unterschiedlichen Basiswahrscheinlichkeiten bei korrektem Verhalten erklärt. Die normierte Erfolgsquote ist in der Success-Bedingung mit 89% höher als in der V- und D-Bedingung (jeweils 81%). Diese Unterschiede erweisen sich in t-Tests jedoch als nicht signifikant. Es kann also konstatiert werden, dass die Versuchspersonen in den drei Bedingungen hinsichtlich des Erreichbaren in etwa gleich erfolgreich entschieden haben.

Der Vergleich der Anzahl korrekter Entscheidungen mit den Resultaten aus Auswertung 3 (Adaptivität des Verhaltens) zeigt, dass Versuchspersonen mit geringer Situations-Adaptivität weniger Erfolg bei der Behandlung der Äffchen haben als adaptiv vorgehende Probanden (im Durchschnitt nur 7.4 Treffer pro Exkursion statt 9.4 Treffer bei adaptivem Verhalten, $p < .001$). Adaptives Verhalten zahlt sich also in diesem Experiment deutlich aus.

8.5 Diskussion

In der Gesamtschau zeigen diese Resultate, dass Validität, Diskriminationsrate und Success durchaus durch Beobachtungslernen erwerbbar sind. Allerdings ist die Aufgabe so schwierig, dass sie längst nicht von allen Versuchspersonen beherrscht wird. (Tabelle 8.9 versucht eine Zusammenfassung der Resultate). Aus den Erfüllungs-Scores kann man herauslesen, dass 21 der 24 Probanden die Diskriminationsrate gut gelernt haben, 18 der 24 Probanden die Validität und 14 der 24 Probanden die Success-Rate. Die Hälfte der Versuchspersonen (12 der 24) lernte alle drei Masse, 6 Personen lernten zwei Masse (viermal V und D und je einmal V und S bzw. D und S). Von den verbleibenden sechs Probanden lernten vier zumindest die Diskriminationsrate und eine die Validität. Lediglich Vp 14 hat kein einziges Mass gelernt (sie ist auch mit nur 19 korrekt behandelten Äffchen die erfolgloseste).

Die Korrelationen der eingeschätzten Cue-Güte mit der tatsächlichen Cue-Güte vermitteln ein ganz ähnliches Bild. Elf der zwölf Personen, die in den Verhaltensmassen allen drei Cue-Güten entsprochen hatten, konnten auch zumindest zwei der drei Ordnungen zufrieden stellend deklarativ angeben. Lediglich Vp 15 fällt hier mit ihren Angaben in der Nachbefragung völlig aus dem Rahmen. Keine einzige Versuchsperson ist bei den deklarativen Angaben deutlich besser als in den Verhaltensdaten.

Abbildung 8.9: Lernfähigkeit der Versuchspersonen, aufgeschlüsselt nach Expeditionen. Zusätzlich zu den Erfüllungsscores innerhalb einer Aufgabe wurde auch dann die Cugüte als gelernt angesehen, wenn sie in die „falsche“ Exkursion exportiert und dort gezeigt wurde. Bei den Einschätzungen aus der Nachbefragung gilt eine Cugüte als gelernt, wenn entweder die Rangkorrelation bestmöglich ist („perfekt“) oder wenn die Pearson-Korrelation hoch ausfällt („> .70“, wobei einmal bei .69 aufgerundet wurde).

VpID	angemessene Erfüllungs-Scores (aus Tabellen 2-4 / 5)			angemessene Einschätzungen (aus Tabelle 6)		
	V	Success	D	V	V*D	D
1	perfekt	> .70	> .70	perfekt	> .70	perfekt
2	perfekt	perfekt	perfekt	perfekt		perfekt
3			perfekt		> .70	perfekt
4	perfekt	perfekt	perfekt	perfekt	> .70	perfekt
5			> .70			
6	> .70	> .70	perfekt	perfekt	> .70	perfekt
7	> .70	> .70	perfekt	perfekt	> .70	perfekt
8	perfekt	perfekt	> .70	perfekt		perfekt
9	> .70	> .70	perfekt	perfekt	> .70	perfekt
10	perfekt	perfekt	> .70	> .70	> .70	perfekt
12			> .70		> .70	perfekt
13		perfekt	> .70		> .70	perfekt
14						
15	> .70	> .70	> .70			
16	perfekt	> .70	> .70	> .70		> .70
17	perfekt	> .70				
18			> .70			perfekt
19	> .70		> .70			> .70
20	> .70			> .70		
21	perfekt		> .70		> .70	> .70
22	> .70	perfekt	> .70	perfekt		perfekt
23	perfekt		perfekt	> .70	> .70	perfekt
24	> .70	> .70	> .70	perfekt	> .70	perfekt
25	> .70		perfekt	perfekt	> .70	perfekt

Mit Bezug auf die drei Gütemasse vermitteln die Befragungsdaten ebenfalls ein recht ähnliches Bild wie die Erfüllungs-Scores: Am leichtesten tun sich die Probanden bei den Diskriminationsraten (19 der 24 können sie mit hoher positiver Korrelation angeben, 16 sogar in der perfekten Rangfolge). Schwieriger gestaltet sich die Validität (14 Personen machen eine zufrieden stellende Angabe, 10 davon in perfekter Rangordnung). Am schwierigsten ist Usefulness, bei der keine Person die exakte Rangordnung trifft, 13 allerdings eine mit den tatsächlichen Werten hoch korrelierende Angabe machen.

Damit zeigen diese Befunde neben der grundsätzlichen Lernbarkeit der Gütemasse, dass Validität und Diskriminationsrate durchaus als separierte Masse erlernt werden und nicht etwa – wie von Newell gemutmasst – bereits als verbundenes Mass des Success unauflöslich wahrgenommen werden (oder als Usefulness, je nachdem, ob die Ratequote bei Nicht-Diskriminieren mit

einbezogen wird oder nicht). Es findet sich im Experiment keine Person, die nur Success beherrscht. Andersherum hingegen gilt, dass von den 14 Probanden, die Success beherrschen, zwölf auch sowohl Validität als auch Diskriminationsrate im Verhalten einsetzen und elf auch numerisch jeweils zufrieden stellend einschätzen können (nur Vp 15, die die Einschätzungen überhaupt nicht hinbekommen hat, schert hier aus). Validität und Diskriminationsrate sind also getrennt vom Success gelernt und, wenn man ein Wirkungsmodell aufstellen wollte, erweisen sie sich eher als notwendige Voraussetzung für das Beherrschen des Success als umgekehrt. (Für Usefulness als vergleichbares Mass ohne Ratequote gilt aus den erfragten Korrelationen analoges.)

Für die Anwendbarkeit der klassischen Suchregel der TTB-Heuristik (absteigende Ordnungen der Validitäten) ist das eine gute Nachricht, denn ohne vom Success separierte Erlernbarkeit der Validitäten wären Menschen nicht in der Lage, TTB im Alltag erfolgversprechend anzuwenden. In Abgrenzung zu kompensatorischen Modellen der Informationsintegration kommt noch der TTB-favorable Befund hinzu, dass die Validitäten, so sie gelernt werden, in der überwiegenden Mehrzahl in der korrekten ordinalen Folge etabliert werden (nur das ist ja für die Anwendung von TTB notwendig). Die beträchtliche durchgängige Unterschätzung der Validitäten (die auch bereits von Newell et al. [2003] und von Hausmann et. al. [2005] berichtet wurde) um einen durchschnittlichen Wert von .13, verbunden mit einem durchschnittlichen richtungsunabhängigen Schätzfehler von .21 stellt bei einem Umgang mit gewichtenden Modellen den möglichen Erfolg einer auf mehreren Cues basierenden Informationsintegration erheblich in Frage. Auch das unterstreicht, dass es vermutlich äusserst nützlich ist, wenn menschliches Denken und Entscheiden über eine Urteilsheuristik wie TTB verfügt.

In ihrer ursprünglichen Form wurde TTB als gedächtnisbasiertes Informationsabfragemodell konzipiert. Da es dort im Regelfall keine nennenswerten Suchkosten gibt, spielt die Anzahl der Versuche bis zum Erreichen des ersten diskriminierenden Cues keine Rolle. Deswegen kann sich die TTB-Suchregel voll und ganz auf die Validität verlassen. Bei einer „externen“ Informationssuche sieht das jedoch anders aus. Hier ist mit begrenzten Ressourcen umzugehen (und sei es nur, dass Informationssuche eine Person davon abhält, andere Ziele zu verwirklichen). Deswegen sind theoretische Überlegungen wie auch Experimente plausibel, welche das Wissen um die Diskriminationsraten von Cues als einen wichtigen Erfolgsfaktor für effizientes Entscheiden berücksichtigen. Gerade die Diskriminationsrate erweist sich in diesem Experiment als besonders leicht erlernbar. In der Argumentationslogik adaptiven Verhaltens lässt die leichte Lernbarkeit (auch und gerade in „kompetitiven“ Situationen mit der Validität) vermuten, dass es für ein solches Konzept auch praktische Anwendungsmöglichkeiten gibt.

Aus den Befunden von Newell et. al. hatte sich die Frage gestellt, ob Menschen eher Success oder eher Usefulness als „natürliches“ integriertes Gütemass betrachten. Die Antworten auf diese Frage fallen auch beim vorliegenden Experiment nicht eindeutig aus. In der Nachbefragung sollte sich der Umstand zunutze gemacht werden, dass Success im Zweialternativenfall grundsätzlich über 50% liegt, Usefulness bei binären Cues mit Diskriminationsraten von zumeist unter .50 jedoch deutlich unter dieser 50%-Marke. Beachtenswert erscheint uns deswegen der Befund aus

Auswertung 5, dass Usefulness so stark überschätzt wurde (im Mittel .46 bei einem tatsächlichen Wert von .31). Denn bei der unterschätzten Validität und der gleichzeitig überschätzten Diskriminationsrate (wie bereits bei Newell et al., 2003) sollte sich $V \times D$ aufgrund der Schätzwerte bei durchschnittlich .32 einpendeln und nicht bei .46, so wie es die Versuchspersonen angegeben haben. Diese massive Überschätzung (die den von Newell berichteten Befunden widerspricht) kann also nur bedeuten, dass die Probanden nicht in der Nachbefragung $V \times D$ „gerechnet“ haben können, denn dann wären sie ja bei einem Wert von etwa .32 gelandet. Vielmehr scheint es so, dass die Befragten (zumindest teilweise) die Ratequote bei Nichtdiskrimination mit einbeziehen (also Usefulness, verschoben in Richtung Success, angeben) und so in die Richtung auf den Wert von .56 kommen, der bei den von ihnen angegebenen Validitäten und Diskriminationsraten der rechnerische Success wäre.

Natürlich entbehrt diese Interpretation nicht einer spekulativen Komponente. Es wäre deswegen interessant, in zukünftiger Forschung Szenarien des Success (nur ein Cue, dann ggf. raten) und der Usefulness (mehrere Cues sequentiell verfügbar) einander als Versuchsbedingungen gegenüberzustellen. Die Entscheidung, sich bei der direkten Befragung nach dem kombinierten Gütemass an die Arbeit von Newell et al. (2003) anzulehnen und nach Nützlichkeit statt nach Success zu fragen, hat zwar wie erwartet zu einem breiteren Datenpool geführt, jedoch auch zu einigen Restriktionen in der unmittelbaren Vergleichbarkeit der Befragungsdaten mit den Verhaltensdaten. Sofern man einen solchen direkten Vergleich in den Mittelpunkt der Resultate stellt, sollte in künftigen Experimenten zum Lernen von Cue-Güte überprüft werden, ob sich nicht auch eine gleich leicht verständliche Abfrage des Success finden lässt.⁴⁴

In der Summe hat es sich für dieses Experiment als fruchtbar erwiesen, in seiner Konstruktion keine Kompromisse an andere, simultan zu beantwortende Fragestellungen machen zu müssen. Die Lernsituation konnte so gestaltet werden, dass sowohl Validität als auch Diskriminationsrate als auch Success bzw. Usefulness durch aktives Beobachten erlernbar waren. Ausserdem wurde den Versuchspersonen keine Hinweise auf eine bestimmte Cue-Ordnung induziert, die sie für das sinnvolle Durchführen eines „Hauptexperimentes“ gebraucht hätten. Schliesslich war die Anwendungsphase nicht darauf ausgerichtet, ein forced choice zwischen gelernten Cue-Rangordnungen zu erzwingen. Vielmehr erwies sich die Möglichkeit, getrennt voneinander auf Verhaltensebene zu demonstrieren, dass man Validitäten, Diskriminationsraten und Success beherrscht, als gute Vorlage, um die Erlernbarkeit der technischen Gütewerte der Cues abschätzen zu können.

⁴⁴ Denkbar wäre beispielsweise: „Angenommen, Du würdest jeden einzelnen Test genau 100-mal hintereinander anwenden. Du würdest Dich auf das Testresultat verlassen, aber müsstest immer dann raten, wenn der Test nicht anspricht. Bei welchem der fünf Tests würdest Du dann am häufigsten, zweit-häufigsten, etc. ein korrektes Ergebnis erzielen?“ Das ist deutlich komplizierter als das einfache „sag mir, wie nützlich der Cue ist“ bei Newell et al., aber es würde einen direkten Vergleich der Verhaltensdaten mit den Befragungsdaten erlauben.

8.6 Literatur

- Attnave, F. (1953). Psychological probability as a function of experienced frequency. *Journal of Experimental Psychology*, 46, 81-86.
- Bröder, A. (2000). Assessing the empirical validity of the "Take The Best"-heuristic as a model of human probabilistic inference. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 26, 1332-1346.
- Bröder, A. (2005). *Entscheiden mit der „Adaptiven Werkzeugkiste“: Ein empirisches Forschungsprogramm*. Lengerich: Pabst Science Publishers.
- Bröder, A., & Schiffer, S. (2003). "Take The Best" versus simultaneous feature matching: Probabilistic inferences from memory and effects of representation format. *Journal of Experimental Psychology: General*, 132, 277-293.
- Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, 34, 571-582.
- Estes, W. K. (1976). The cognitive side of probability learning. *Psychological Review*, 83, 37-64.
- Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, 103, 650-669.
- Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic Mental Models: A Brunswikan theory of confidence. *Psychological Review*, 98, 506-528.
- Hasher, L. & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General*, 108, 356-388.
- Hasher, L. & Zacks, R. T. (1984). Automatic processing of fundamental information. *American Psychologist*, 39, 1372-1388.
- Hausmann, D., Christen, S. & Läge, D. (2005). *Und sie lernen es doch! Das Einschätzen der Validität von probabilistischen Hinweiscues und der Einfluss der subjektiven Repräsentation auf das Urteilsverhalten*. AKZ-Forschungsbericht 08. Zürich: Angewandte Kognitionspsychologie.
- Läge, D., Hausmann, D., Christen, S. & Daub, S. (2005). *Was macht einen „guten Cue“ aus? Strategien der Informationssuche beim heuristischen Entscheiden unter Unsicherheit*. AKZ-Forschungsbericht 05. Zürich: Angewandte Kognitionspsychologie.
- Lee, M. D., & Cummins, T. D. R. (2004). Evidence accumulation in decision making: Unifying the "take the best" and the "rational" models. *Psychonomic Bulletin and Review*, 11, 343-352.
- Newell, B. R., & Shanks, D. R. (2003). Take the best or look at the rest? Factors influencing "one-reason" decision-making. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 29, 53-65.
- Newell, B. R., Rakow, T., Weston, N. J., & Shanks, D. R. (2003). Search strategies in decision-making: the success of "success". *Journal of Behavioral Decision Making*, 17, 117-137.
- Newell, B. R., Weston, N. J., & Shanks, D. R. (2003). Empirical tests of a fast and frugal heuristic: not everyone "takes-the-best". *Organizational Behavior and Human Decision Processes*, 91, 82-96.
- Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. Cambridge, MA: Cambridge University Press.
- Peterson, C. R. & Beach, L. R. (1967). Man as an intuitive statistician. *Psychological Bulletin*, 68, 29-46.
- Shuford, E. H. J. (1961). Percentage estimation of proportion as a function of element type, exposure type, and task. *Journal of Experimental Psychology*, 61, 430-436.

Teil IV - Trinkwasser- &-Börsen-Szenario

Kapitel 9

Trinkwasser- und Börsenszenario: Aufbau, Ablauf und technische Spezifikation einer Experimentalumgebung zur entscheidungsvorbereitenden Informationssuche in zeitlichen Verläufen

In Alltagsentscheidungen spielen Informationen eine wichtige Rolle. Diese sind aber nicht immer top-aktuell, sondern zum Teil bereits älteren Datums, was eine Entscheidung auf dieser Basis unsicher macht. Je geringer die zeitliche Stabilität eines Prädiktors und einer Zielvariable also sind, umso weniger zuverlässig die Information. Dieser fundamentale Zusammenhang wurde in ein Experimentaldesign eingebaut, welches sich – in Anlehnung an bereits bestehende Untersuchungen – an einem Börsenszenario mit Zweialternativenfall orientiert. Hier lassen sich die zeitliche Stabilität von Cue und Zielvariablen sehr gut als zusätzliche Elemente realisieren.

Mit der Einführung der Zeitdimension in die Entscheidungsforschung taucht jedoch auch eine neue Ressource auf, die nicht einfach mit vergleichbar ist: Zeit selbst nämlich. Sie vergeht, und man kann sie nicht mehr. Um dem Rechnung zu tragen, wurde mit dem Trinkwasser-Szenario eine neue Experimentalidee entwickelt, in der Zeit zur zentralen Ressource wird.

Diese beiden Experimente – sie sind bezüglich ihrer statistischen Umwelten prinzipiell gleich aufgebaut und damit in den Resultaten miteinander vergleichbar – werden im vorliegenden Forschungsbericht in ihrem grundsätzlichen dreistufigen technischen Aufbau mit allen Details vorgestellt. Dieser Forschungsbericht ergänzt damit weitere Arbeiten, die sich mit den einzelnen psychologischen Fragestellungen auseinandersetzen und diese mit der Anwendung eben dieser Experimentaldesigns zu beantworten suchen.

9.1 Einleitung

In bisherigen Experimenten zur ecological rationality hat sich klar gezeigt, dass die Informationssuche ein zentraler Faktor in Entscheidungsprozessen ist. Die Informationssuche wurde bis anhin nur bei unabhängigen Einzelereignissen, d.h. ohne Einbezug einer Zeitdimension untersucht und fand jeweils innerhalb einer stabilen Umwelt statt. Mit den beiden hier vorgestellten Experimentaldesigns fokussieren wir nun auf die Anbindung probabilistischer mentaler Modelle (nach Gigerenzer et al., 1999) an sich verändernde Umwelten und dynamische Cue-Landschaften.

Die für diese Designs entscheidende Neuerung gegenüber bisherigen Experimentalanordnungen im Bereich der Informationssuche ist die, dass die Versuchspersonen das gesamte Experiment nicht als eine Reihe von Einzeldurchgängen erleben, sondern in einen Zeitfluss eintauchen. Sämtliche entscheidungsrelevanten Informationen werden dabei auf einem information board (vgl. Payne 1976) dargeboten. Gemessen wird das individuelle Entscheidungsverhalten anhand der Suchrichtung, dem Abbruchkriterium und der Entscheidungsregel (vgl. Payne, Bettman & Johnson, 1993).

Wir wollen wissen, ob Personen bei Entscheidungen, die in einem Zeitfluss miteinander verbunden sind, die zeitliche Dimension und die Informationen über die zeitliche Stabilität eines Systems in der Informationsnutzung berücksichtigen (d.h. Informationen aus der Vergangenheit berücksichtigen, gleichzeitig aber auch auf ihre „Verfallsquote“ achten). Ausserdem ist interessant, wie weit Personen Informationen nicht nur für die Gegenwart suchen, sondern auch den prospektiven, also den in die Zukunft gerichteten Nutzen einer Information in ihr Suchverhalten einbeziehen.

Als Informationen über die zeitliche Stabilität des Systems führen wir zwei unterschiedliche strukturelle Parameter ein: Erstens SZ (die Stabilitätsrate einer Entscheidungs-Alternative Z, kurz „Zielvariablen-Stabilität“) und zweitens SI (die Stabilitätsrate einer Informationsquelle I, kurz „Cue-Stabilität“).

Neben der Validität eines Cues und seiner Diskriminationsrate (die in diesem Experimentaldesign vernachlässigt wird) führen wir also die Cue-Stabilitätsrate als ein drittes Cue-Gütemass ein: Ein zeitlich weniger stabiler (aber hoch valider) Cue liefert Informationen, auf die man sich schon nach kurzer Zeit nicht mehr verlassen kann, während die Information eines zeitlich stabilen (und hoch validen) Cues auch nach mehreren (im Experimentaldesign ca. zwei bis drei) Zeitschritten noch immer eine relativ hohe Validität aufweist.

Für die Zielvariablen-Stabilität gilt, dass die Ausprägung einer zeitlich stabileren Zielvariable über einen längeren Zeitraum hinweg konstant bleibt – was bei einer zeitlich weniger stabilen Zielvariable nicht der Fall ist.

Die Stabilitätsrate definieren wir allgemein als Wahrscheinlichkeit, mit der die Ausprägung einer Variablen im nächsten Durchgang (bzw. zum nächsten Messzeitpunkt) noch dieselbe ist wie im

aktuellen Durchgang (bzw. zum aktuellen Messzeitpunkt). Eine zeitlich unveränderliche, völlig stabile Ausprägung (einer Zielvariablen oder eines Cues) weist somit eine Stabilitätsrate von 100% auf. Eine Variable, deren Ausprägung sich von Durchgang zu Durchgang (bzw. von Messzeitpunkt zu Messzeitpunkt) ändert, hat hingegen eine Stabilitätsrate von 0%. Die Stabilitätsrate ist also abhängig von der Anzahl der festgelegten Messzeitpunkte und daher kein absoluter Wert. Sie wird angegeben unter der Massgabe von Beobachtungszeiträumen (d.h. als prozentualer Anteilswert).

Welchen Einfluss auf das Such- und Entscheidungsverhalten im Zwei-Alternativenfall hat nun einerseits das Wissen um die Stabilitäten der Zielvariablen und andererseits das Wissen um die Stabilitäten der Cues? Und welchen Einfluss hat die Verwendung unterschiedlicher Ressourcen, genauer: Was passiert, wenn wir die üblicherweise verwendeten monetären Ressourcen durch die Ressource Zeit ersetzen? Da Zeit ein wichtiger Faktor in Entscheidungssituation ist (vgl. Maule & Edland 1997, S. 189), soll das Experiment auch zeigen, ob in Abhängigkeit der verwendeten Ressource Unterschiede im Such- und/oder Entscheidungsverhalten auftreten.

Um diese Fragen zu untersuchen, haben wir uns zur Durchführung eines explorativen Experiments entschieden, dessen Aufbau, Ablauf, Designüberlegungen und technische Spezifikation in den nachfolgenden Kapiteln beschrieben ist. Das Experiment wurde in Microsoft Visual Basic 6.0 programmiert und vollständig auf einem PC mit Windows XP-Betriebssystem durchgeführt. Eine Beschreibung der detaillierten Fragestellungen und der Ergebnisse findet sich in Zurbriggen, Hausmann, Christen & Läge (2007), Zurbriggen, Christen, Hausmann & Läge (2007a und 2007b), sowie Zurbriggen & Läge (2007).

9.2 Experimentalaufbau

9.2.1 Allgemeine Struktur des Experiments

Aufteilung in zwei Teilexperimente. Primär soll das Design erkunden können, wie sich die Einführung der Zeitdimension sowie der Zielvariablen- und der Cue-Stabilität auf das Informationssuch- und Entscheidungsverhalten auswirkt. Um auch die Auswirkungen unterschiedlicher Ressourcen (Zeit vs. Geld) untersuchen zu können, ist das Experiment in zwei zusammenhängende, miteinander vergleichbare Teilexperimente aufgeteilt, in denen je eine der beiden Ressourcen verwendet wird (Experiment 1A: Ressource = Zeit; Experiment 1B: Ressource = Geld). Den Einfluss der Zeitdimension und der Zielvariablen-Stabilität soll mit Experiment 1A klären. Experiment 1B dient grundsätzlich der Untersuchung derselben Fragestellungen wie Experiment 1A, wobei hier zusätzlich der Einfluss der Cue-Stabilität untersuchbar sein soll.

Kontextuelle Umwelt. Da in den beiden Experimenten je unterschiedliche Ressourcen verwendet werden, unterscheiden sich diese hinsichtlich ihrer kontextuellen Umwelt (d.h. hinsichtlich der Einbettung der sich wiederholenden Entscheidungsaufgabe in eine konstruierte Entscheidungssituation). Bei Experiment 1A sprechen wir aufgrund der gewählten Situationseinbettung auch von einem „Trinkwasser-Szenario“ und bei Experiment 1B von einem „Börsenszenario“. Das Börsenszenario ist angelehnt an das Börsenexperiment von Christen, Hausmann & Läge (2007),

wobei im vorliegenden Experiment die Informationen nur einzeln abrufbar sind und die Diskriminationsrate vernachlässigt wird. Das Trinkwasser-Szenario hingegen ist völlig neu entwickelt.

Statistische Umwelt. Bezüglich ihrer statistischen Umwelten sind die beiden Experimente prinzipiell identisch aufgebaut: Die Entscheidungssituation ist innerhalb des Theorierahmens des One-Reason Decision Making (ORDM) formuliert. Das Verhältnis zwischen den Kosten für einen einzelnen Cue und dem maximal erzielbaren Rundengewinn in allen Bedingungen beträgt 1:10. Dieses nimmt jedoch in den beiden Experimenten unterschiedliche Formen an. Im Trinkwasser-Szenario kostet eine Information Zeit und im Börsenszenario Geld. Zeit kann zwar nicht hinzugewonnen oder vervielfacht werden (wie die Geldbeträge im Börsenszenario), dafür kann sie unterschiedlich verwendet werden. (Details zur statistischen Umwelt können den folgenden Abschnitten und Kapiteln entnommen werden. Festgehalten sei an dieser Stelle, dass die beiden Experimente aufgrund der kongruent gewählten statistischen Umwelten direkt miteinander bezüglich der Verwendung unterschiedlicher Ressourcen verglichen werden können).

Experimentalbedingungen. Beide Experimente sind in je drei Experimentalbedingungen unterteilt: Das Trinkwasser-Szenario (mitunter auch als „Quellen-Szenario“ betitelt) in die Bedingungen Q1, Q2 und Q3 und das Börsenszenario in die Bedingungen B1, B3 und B4. Die Nummerierung ist so gewählt, dass Bedingungen mit identischen statistischen Umwelten sowohl im Trinkwasser- als auch im Börsenszenario dieselbe Nummer tragen. Bezüglich ressourcenabhängiger Unterschiede kann deshalb Q1 direkt mit B1, und Q3 direkt mit B3 verglichen werden. Eine Bedingung B2 existiert nicht, weil keine der Börsenszenario-Bedingungen dieselbe statistische Umwelt wie Q2 besitzt. Dasselbe gilt, in analoger Weise für das Nichtvorhandensein einer Bedingung Q4.

Q1 und B1 sind die Basisbedingungen der beiden Experimente. Mit ihnen erfassen wir das Entscheidungsverhalten bei voneinander unabhängigen Einzelereignissen (Entscheidungsaufgaben). Dementsprechend gibt es in Q1 und B1 noch keine Zeitdimension und keine Angaben zu den Cue- und den Zielvariablen-Stabilitäten. Bei den übrigen Bedingungen (Q2, Q3, B3, B4) sind die Ereignisse seriell über die Zeitachse miteinander verbunden und wir geben in allen Bedingungen (ausser Q2) die Zielvariablen-Stabilitäten bekannt. In B3 sind die Stabilitäten beider Zielvariablen identisch, in Q3 hingegen unterschiedlich. Im Börsenszenario gibt es eine zusätzliche Bedingung B4. B4 unterscheidet sich von B3 darin, dass wir den Versuchspersonen in B4 nicht nur die Stabilitäten der Zielvariablen sondern auch die Stabilitäten der Cues bekannt geben.

Das Trinkwasser-Szenario: In den drei Bedingungen von Experiment 1A besteht die Aufgabe der Versuchspersonen darin, in Tanklastwagen abgefülltes Quellwasser mittels Wasserqualitäts-tests (= Cues) zu analysieren und zu entscheiden, ob das Wasser für die Abfüllung als Trinkwasser geeignet ist oder ob es weggeschüttet werden soll. Pro Runde stehen immer zwei Tankladungen zur Auswahl. Die beiden Tankladungen stammen entweder aus zwei unbekannten, sich nicht systematisch wiederholenden Wasserquellen (Q1) oder immer aus denselben beiden Quellen, d.h. jeweils die eine Tankladung aus Quelle X und die andere aus Quelle Y (Q2, Q3). In allen drei Bedingungen sind die Wasserquellen jeweils voneinander unabhängig. Aus Kapazitätsgrün-

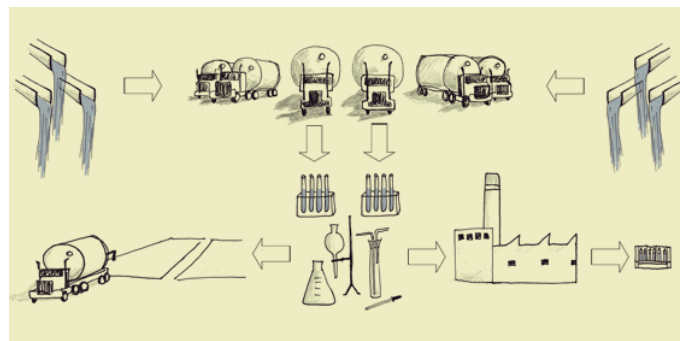
den kann maximal eine der beiden Tankladungen zu Trinkwasser abgefüllt werden. Es ist aber auch erlaubt, beide Ladungen wegzuschütten. Erst nach der definitiven Entscheidung und nach dem Abfüllen in Trinkwasserflaschen wird durch eine staatliche Kontrolle ermittelt, ob das Wasser tatsächlich zum Verkauf freigegeben wird oder nicht.

Trinkwasser-Produktion – Anleitung

Du bist für 3 Monate PraktikantIn in der kleinen Firma AguaMex in Mexiko. AguaMex füllt Quellwasser in Trinkwasserflaschen ab und verkauft diese anschliessend an Läden.

Das Quellwasser wird jeweils von verschiedenen Wasserquellen geholt und mit Tanklastwagen zu AguaMex geliefert, wo es eine Wasseranalysestation und eine Abfüllstation gibt.

Bevor das abgefüllte Trinkwasser jedoch verkauft werden kann, muss es noch von der staatlichen Lebensmittelkontrollstelle für gut befunden werden. Diese Kontrolle ist deshalb notwendig, weil die Qualität des Quellwassers kritisch ist (nur rund ein Drittel aller Tankladungen sind überhaupt für die Trinkwasserherstellung geeignet). Wenn die Qualität von der Kontrollstelle als schlecht eingestuft wird, waren das Abfüllen und die von dir investierte Zeit umsonst.



Es kommen jeweils zwei Lastwagen mit Wasser von zwei verschiedenen Quellen gleichzeitig bei AguaMex an.

In der Wasseranalysestation hast du die Möglichkeit, bis zu vier Qualitätstests pro Tankladung durchzuführen. Es steht dir frei, ob du Tests durchführen willst, wie viele, welche und bei welcher Tankladung. Falls du die Wasserqualität einer Tankladung für gut befindest, kannst du mit dieser Tankladung zur Abfüllstation fahren.

Aus Kapazitätsgründen musst du dich für eine der zwei Tankladungen entscheiden. Den anderen Lastwagen schickst du zu den benachbarten Landwirten, die damit ihre Felder bewässern. Falls du der Meinung bist, dass keine der beiden Tankladungen trinkbares Quellwasser enthält, kannst du beide Lastwagen zu den benachbarten Landwirten schicken. Dies bringt dir mit Sicherheit keine Einnahmen, dafür verbrauchst du jedoch keine Abfüllzeit.

Abbildung 9.1: Allgemeiner Einleitungstext zum Trinkwasserszenario

Die Wasserqualitätstests können je einzeln pro Tankladung durchgeführt werden. Die Anwendung eines Tests kostet eine Stunde Zeit, die von einem Gesamtzeitbudget von 240 Stunden abgezogen wird. Maximal können bei beiden Tankladungen jeweils vier Tests, also insgesamt acht, durchgeführt werden, sofern noch genügend Gesamtzeit zur Verfügung steht. Das Anwenden einer minimalen Anzahl an Tests wird nicht gefordert, d.h. es ist erlaubt, sich ohne die Anwendung auch nur eines einzigen Tests für eine der beiden Alternativen (oder für das Wegschütten

beider Tankladungen) zu entscheiden. Ein Abfüllvorgang in Trinkwasserflaschen beansprucht immer zehn Stunden Zeit. Auch diese wird vom Gesamtzeitbudget abgezogen. Das Spiel ist beendet, wenn das Zeitbudget aufgebraucht ist. Der den Versuchspersonen präsentierte allgemeine Einleitungstext zum Trinkwasser-Szenario ist in Abbildung 9.1 im Wortlaut aufgeführt.

Das Börsenszenario. In den drei Bedingungen von Experiment 1B werden den Versuchspersonen an jedem Handelstag zwei alternative Optionsscheine (Q1) bzw. Fonds (Q3, Q4) vorgelegt. Die Kursverläufe der beiden Alternativen sind gegenseitig unabhängig. Beide Alternativen können im Wert entweder steigen oder sinken, was durch einen diagonal nach oben bzw. unten zeigenden Pfeil dargestellt wird. In jeder Runde kann auf höchstens eine der beiden Alternativen gesetzt werden. Pro Runde steht ein Spielgeld von genau CHF 10'000.- zur Verfügung. Mit Hilfe von bis zu vier Kurs-Indikatoren (Cues) sollen die Versuchspersonen entscheiden, ob sie in eine der beiden Alternativen investieren wollen oder in keine von beiden. Die Indikatoren können je einzeln pro Alternative abgefragt werden und ihr Einsatz kostet je CHF 1'000.- (= Informationssuchkosten). Die Informationssuchkosten betragen somit maximal CHF 8'000.- und werden vom Spielgeld abgezogen. Es ist erlaubt, sich ohne die Abfrage von Indikatoren für eine der beiden Alternativen zu entscheiden. Das Spielgeld, das nach der Informationssuche noch übrig ist, kann entweder auf eine der beiden Alternativen gesetzt oder direkt auf das Gewinnkonto transferiert werden. Das Setzen auf eine der beiden Alternativen resultiert im Erfolgsfall, d.h. bei steigendem Kursverlauf der ausgewählten Alternative, in einer Gewinnabschöpfung in dreifacher Höhe des Einsatzes. Das gewonnene Geld wird sodann dem Gewinnkonto gutgeschrieben. Im Misserfolgsfall resultiert anstelle eines Gewinns der Verlust des Einsatzes, und das Gewinnkonto erfährt keinen Zuwachs. Nach 20 Handelstagen endet das Spiel.

Dein Zeitbudget während deines Praktikums beträgt 240 Arbeitsstunden. Um eine einzelne Tankladung abzufüllen, brauchst du 10 Stunden deiner Arbeitszeit. Jeder Test, den du durchführst, dauert 1 Stunde.

Dein Ziel besteht darin, in der verfügbaren Zeit möglichst viele Ladungen von gutem Trinkwasser abzufüllen.

Die staatliche Kontrollstelle teilt dir in unregelmässigen Abständen mit, wie viele deiner Trinkwasserladungen überhaupt gut genug und verkäuflich waren.

Du bekommst noch einige Angaben von deiner Praktikums-Vorgängerin zu den Tests zur Analyse der Wasserqualität:

Jeder der 4 Tests gibt dir Hinweise, ob das Quellwasser gut (✓) oder nicht gut (X) ist.

Deine Vorgängerin hat langfristige Erfahrung über die Verlässlichkeit der Tests gemacht. Die Verlässlichkeit dieser Tests wird durch die Trefferquoten ausgedrückt. Die Trefferquote, eine Zahl zwischen 0 und 100, gibt also an, in wie vielen von 100 Fällen ein Test nach ihrer Erfahrung eine korrekte Vorhersage macht. Bei einer Trefferquote von 100 sagt ein Test mit Sicherheit korrekt voraus, ob die Tankladung trinkbares Wasser enthält. Bei einer Trefferquote von 50 sagt der Test in 50 von 100 Fällen das Richtige und in den anderen 50 Fällen das Falsche voraus.

Drücke auf "Start", um das Experiment zu starten !

Abbildung 9.2: Instruktion zu Bedingung Q1

9.2.2 Experimentalbedingungen des Trinkwasser-Szenarios im Detail

Quellenbedingung 1 (Q1). Q1 ist die Basisbedingung für Experiment 1A. Sie dient der Erhebung einer Baseline. In Q1 stehen in jeder Runde jeweils zwei neue Alternativen zur Auswahl, von deren Herkunft man nichts weiss. Diese Alternativen stehen somit in keinem Zusammenhang zueinander und sind zeitlich nicht miteinander verknüpft. Eingebettet in den Kontext des Trinkwasser-Szenarios bedeutet dies, dass die Ladungen der beiden Tanklastwagen in jeder Runde von zwei potenziell anderen Wasserquellen stammen. Das information board besteht aus einer 2x4-Matrix (2 Alternativen x 4 Cues). Darauf ersichtlich sind die Validitäten der Cues sowie die Schaltflächen für den Informationsabruf. Die Instruktion zu Bedingung Q1 ist in Abbildung 9.2 im Wortlaut ersichtlich. Abbildung 9.3 zeigt die Graphische Oberfläche von Q1.

Analyse	Trefferrate	Quelle X	Quelle Y
Test A	75	?	?
Test B	80	?	?
Test C	90	✗	✓
Test D	85	?	✗

Legende : ✗ = schlechte Wasserqualität ✓ = gute Wasserqualität

Handlung

Wasser aus Quelle X

Wasser aus Quelle Y

Keine Abfüllung

→

→

→

Auswahl bestätigen

Übersicht

Aktueller Zeitverbrauch (in Stunden) : 13

Noch vorhandene Zeit (in Stunden) : 215

Anzahl brauchbare Lieferungen bisher : 0

Von den letzten 1 Lieferungen brauchbar : 0

Anzahl brauchbare Lieferungen insgesamt : 0

Nächste Lieferungen

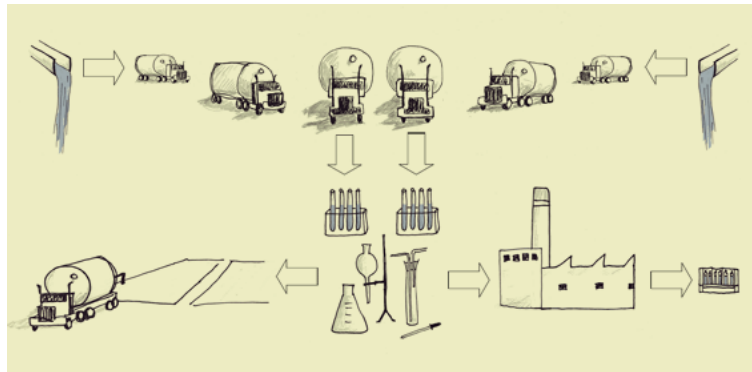
Abbildung 9.3: Graphische Oberfläche von Versuchsbedingung Q1

Quellenbedingung 2 (Q2). Bedingung Q2 unterscheidet sich von Q1 erstens darin, dass in Q2 immer dieselben beiden Alternativen (Tankladungen aus den Wasserquellen X und Y) zur Verfügung stehen. Zweitens bleiben die Cue-Informationen der letzten vier Runden auf dem Information Board sichtbar. Dadurch entsteht eine Zeitachse, auf der der Verlauf der bisher durchgeführten Tests sichtbar ist. Damit wird nicht nur die Konstanz der Entscheidungsalternativen und die zeitliche Abhängigkeit der Ereignisse unterstrichen, sondern die Versuchsperson erhält gleichzeitig die Möglichkeit, vorausgegangene Testergebnisse zu nutzen (und das, ohne ihr Gedächtnis bemühen zu müssen). Die Instruktion zu Bedingung Q2 ist in Abbildung 9.4 im Wortlaut aufgeführt. Abbildung 9.5 zeigt die Benutzeroberfläche von Q2.

Trinkwasser-Produktion (Fortsetzung) – Anleitung

Da es dir in Mexiko so gut gefallen hat, entscheidest du dich spontan, nochmals 3 Monate für Aguamex zu arbeiten.

In deiner Arbeit musst du dich jedoch ein wenig umstellen, denn es stehen für die Anlieferung **nur noch zwei Wasserquellen** zur Verfügung. Da die Wasserquellen immer dieselben zwei sind und sich die Wasserqualität nicht willkürlich ändert, bleiben die **Informationen der Tests der letzten vier Zeitpunkte** nutzbar. Die ankommenden Tanklastwagen bilden zwei Schlangen, die eine von Wasserquelle X kommend, die andere von Wasserquelle Y kommend.



Du musst dich jeweils entscheiden, ob du eine der beiden Ladungen abfüllen willst und falls ja, welche.

Drücke auf "Start", um mit der Fortsetzung des Experiments zu beginnen !

Abbildung 9.4: Instruktion zu Bedingung Q2

Analyse	Trefferrate	Quelle X	Quelle Y
Test A	75	?	?
Test B	80	?	?
Test C	90	✗	✓
Test D	85	?	✗

Legende : ✗ = schlechte Wasserqualität ✓ = gute Wasserqualität

Handlung

Wasser aus Quelle X

Wasser aus Quelle Y

Keine Abfüllung

Auswahl bestätigen

Übersicht

Aktueller Zeitverbrauch (in Stunden) : 13

Noch vorhandene Zeit (in Stunden) : 215

Anzahl brauchbare Lieferungen bisher : 0

Von den letzten 1 Lieferungen brauchbar : 0

Anzahl brauchbare Lieferungen insgesamt : 0

Nächste Lieferungen

Abbildung 9.5: Graphische Oberfläche von Versuchsbedingung Q2

Quellenbedingung 3 (Q3). Bedingung Q3 ist, mit einer Ausnahme, identisch mit Bedingung Q2: Auch in Q3 ist die Zeitachse mit den Cue-Informationen der letzten vier Runden sichtbar, jedoch sind auf dem information board zusätzlich die Stabilitäten der beiden Alternativen explizit angegeben (Alternative X: $S=.60$; Alternative Y: $S=.90$). Damit wird den Versuchspersonen mitgeteilt, wie häufig sich die Wasserqualität einer Quelle im Mittel ändert (von trinkbar zu unge-niessbar oder umgekehrt). So lässt sich untersuchen, ob die Kenntnis der Stabilitäten der Zielva-riablen das Informationssuch- und Entscheidungsverhalten der Versuchspersonen beeinflusst. Die Instruktion zu Q3 ist in Abbildung 9.6 wiedergegeben. Abbildung 9.7 zeigt die Benutzeroberfläche von Q3.

Trinkwasser-Produktion (Fortsetzung)

Da es dir in Mexiko immer noch sehr gut gefällt, entscheidest du dich, noch ein letztes Mal während 2 Monaten für AguaMex zu arbeiten.

Verglichen mit den letzten zwei Monaten gibt es **eine einzige Änderung**:

Du weisst nun auch etwas über die **Stabilität der Wasserqualität der beiden Quellen**.

Die Stabilität der Wasserqualität (eine Zahl zwischen 0 und 100) gibt an, in wie vielen von 100 Fällen die Wasserqualität bei der nächsten Messung noch dieselbe ist. Beträgt die Stabilität 100, ändert sich die Wasserqualität nie. Beträgt die Stabilität 50, so ist die aktuelle Wasserqualität in der Hälfte der Fälle dieselbe wie bei der letzten Messung.

Drücke auf "Start", um mit der Fortsetzung des Experiments zu beginnen!

Abbildung 9.6: Instruktion zu Bedingung Q3

Analyse	Trefferrate	Quelle X mit Stabilität = 60					Quelle Y mit Stabilität = 90				
Test A	75	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Test B	80	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Test C	90	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Test D	85	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Vergangene Tage seit Erhalt der Info : 4 3 2 1 0 4 3 2 1 0

Legende : ☒ = schlechte Wasserqualität ☒ = gute Wasserqualität

Handlung

Wasser aus Quelle X ☐

Wasser aus Quelle Y ☒

Keine Abfüllung ☐

Auswahl bestätigen

Übersicht

Aktueller Zeitverbrauch (in Stunden) : 10

Noch vorhandene Zeit (in Stunden) : 192

Anzahl brauchbare Lieferungen bisher : 0

Von den letzten ? Lieferungen brauchbar : ?

Anzahl brauchbare Lieferungen insgesamt : ?

Nächste Lieferungen

Abbildung 9.7: Graphische Oberfläche von Versuchsbedingung Q3

Optionsschein-Handel – Anleitung

Du bist TeilnehmerIn an einem vereinfachten Börsenspiel, in dem du an insgesamt **20 Handelstagen** mit verschiedenen Optionsscheinen handelst.

An jedem Handelstag steht dir ein Budget von CHF **10'000.-** zur Verfügung. Diesen Betrag kannst du jeweils auf einen von zwei Optionsscheinen setzen. Nach der Durchführung des Handels siehst du nach einigen Handelstagen, ob die gewählten Optionsscheine Gewinn abgeworfen haben oder nicht. Insgesamt wird **rund ein Drittel** aller Optionsscheine **gewinnen** und die anderen zwei Drittel werden verlieren.

Falls du auf einen Optionsschein setzt, der an Wert gewinnt, wird dir **der 3-fache Betrag deines Einsatzes** auf dein Konto ausbezahlt. Verliert der Optionsschein jedoch an Wert, ist der eingesetzte Betrag vollständig verloren. Falls du weder auf den einen noch auf den anderen Optionsschein setzen willst, kannst du den einsetzbaren Betrag statt dessen in die **Sparbüchse** legen (d.h. er wird einfach deinem Konto gutgeschrieben).

Dein Ziel besteht darin, einen möglichst hohen Kontostand zu erreichen.

Es besteht die Möglichkeit zu jedem der Optionsscheine Informationen zu kaufen, welche angeben mit welcher Wahrscheinlichkeit ein Optionsschein gewinnen oder verlieren wird: **4 Indikatoren** (Informationsquellen) geben dir entweder mit einem steigenden Pfeil (= Optionsschein wird an Wert gewinnen), oder einem sinkenden Pfeil (= Optionsschein wird an Wert verlieren) Hinweise.

Die Verlässlichkeit dieser Indikatoren wird durch die **Trefferquoten** ausgedrückt. Die Trefferquote (eine Zahl zwischen 0 und 100) gibt also an, in wie vielen von 100 Fällen ein Indikator eine korrekte Vorhersage macht. Beträgt die Trefferquote 100, so ist die von dem Indikator gemachte Vorhersage in jedem Fall richtig. Bei einer Trefferquote von 50 ist die Information in 50 von 100 Fällen richtig und in den restlichen 50 Fällen falsch.

Das Aufdecken von Informationen ist mit Kosten verbunden. **Jede Information kostet dich CHF 1'000.-**. Diese werden jeweils von deinem Budget abgezogen und du kannst sie weder auf einen Optionsschein setzen, noch in die Sparbüchse tun.

Melde dich bitte bei der Versuchsleitung,

Abbildung 9.8: *Instruktion zu Bedingung B1*

9.2.3 Experimentalbedingungen des Börsenszenarios im Detail

Börsenbedingung 1 (B1): Bedingung B1 ist die Basisbedingung. Sie dient zur Erhebung einer baseline für Experiment 1B. In Bedingung B1 werden in jeder Runde jeweils zwei völlig neue Alternativen (Optionsscheine) angeboten. Die Versuchspersonen können das nach der Informationssuche verbliebene Runden-Spiel-geld entweder in einen der beiden Optionsscheine investieren oder direkt auf das Gewinnkonto transferieren. Das information board besteht (wie schon in Q1) aus einer 2x4-Matrix (2 Alternativen x 4 Cues). Darauf ersichtlich sind die Validitäten der Cues sowie die Schaltflächen für den Informationsabruf. Die Instruktion zu Bedingung B1 ist in Abbildung 9.8 im Wortlaut ersichtlich. Abbildung 9.9 zeigt die Benutzeroberfläche von B1.

Börsenbedingung 3 (B3): Im Unterschied zu B1 stehen in B3 immer dieselben beiden Alternativen zur Auswahl. Und so wie in Q3 bleibt nun auch in B3 der zeitliche Verlauf der Informationssuche (die während der letzten vier Runden abgerufenen Cue-Informationen) sichtbar. Ebenso wie in Q3 werden auch hier die Stabilitäten der beiden Alternativen explizit angegeben. Im Gegensatz zu B3, wo die Stabilitäten beider Alternativen unterschiedlich sind, sind sie hier mit

$S=.80$ jedoch gleich gross. Die Einbettung in den Kontext des Börsenszenarios legt nahe, die beiden Alternativen in B3 aus Plausibilitätsgründen nicht mehr als Optionsscheine, sondern als Fonds zu bezeichnen (weil immer das gleiche Alternativenpaar zur Auswahl steht, was für Optionsscheine keinen Sinn machen würde). Die Instruktion zu Bedingung B3 ist in Abbildung 9.10 aufgeführt. Abbildung 9.11 zeigt die Benutzeroberfläche von B3.

Informationen	Trefferrate	Optionsschein X	Optionsschein Y
Indikator A	80	?	?
Indikator B	75	?	?
Indikator C	90	↑	?
Indikator D	85	?	?

Handel
Budget 10'000 minus Suchkosten 1'000 = investierbarer Betrag 9'000 ? → Optionsschein X 9'000 ? → Optionsschein Y 0 ? → Sparbüchse 0 Auswahl bestätigen

Übersicht
Bisheriger Kontostand : 8'000 Gewinn in den letzten 5 Tagen : 116'000 (darunter 0 Tage ohne Gewinn) Neuer Kontostand : 124'000 Anzahl vergangener Handelstage : 5 Nächster Handelstag

Abbildung 9.9: Graphische Oberfläche von Versuchsbedingung B1

Ab nun handelst du anstelle von verschiedenen Optionsscheinen mit nur noch **zwei Fonds** (Fonds sind aus mehreren Aktien, Obligationen, etc. zusammengesetzt). Jeden Tag kannst du erneut bis zu 10'000.- auf einen der beiden Fonds setzen.

Da es sich immer um **dieselben zwei Fonds** handelt, bleiben die von dir aufgedeckten Informationen der **letzten vier Handelstage** sichtbar. Neben den aktuellen Informationen stehen dir also auch **zurückliegende Informationen** zur Verfügung.

Zusätzlich erhältst du **Informationen über die Stabilität der beiden Fonds**. Eine Stabilität von 100 bedeutet, dass der Fonds an jedem Tag denselben Trend („Wertgewinn“ / „Wertverlust“) hat, wie am Vortag. Eine Stabilität von 50 bedeutet, dass der Fonds in der Hälfte der Tage denselben Trend hat, wie am Vortag.

Abbildung 9.10: Instruktion zu Bedingung B3

Börsenbedingung 4 (B4): Bedingung B4 unterscheidet sich von Bedingung B3 lediglich darin, dass in B4 zusätzlich die Stabilitäten der Cues explizit auf dem information board angegeben

sind. So lässt sich untersuchen, ob und wie die zusätzliche Kenntnis der Stabilitäten der Cues das Informationssuch- und Entscheidungsverhalten der Versuchspersonen beeinflusst. Die Instruktion zu Bedingung B4 ist in Abbildung 9.12 wiedergegeben. Abbildung 9.13 zeigt die Benutzeroberfläche von B4.

Informationen	Trefferrate	Fonds X mit Stabilität = 80	Fonds Y mit Stabilität = 80
Indikator A	80	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> ?	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> ?
Indikator B	75	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> ?	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> ?
Indikator C	90	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> ?	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> ?
Indikator D	85	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> ?	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> ?
Vergangene Tage seit Erhalt der Info :		4 3 2 1 0	4 3 2 1 0

Handel
Budget 10'000 minus Suchkosten 1'000 = investierbarer Betrag 9'000 ? → Fonds X 9'000 ? → Fonds Y 0 ? → Sparbüchse 0 Auswahl bestätigen

Übersicht
Bisheriger Kontostand : 0 Gewinn in den letzten 4 Tagen : 59'000 (darunter 1 Tage ohne Gewinn) Neuer Kontostand : 59'000 Anzahl vergangener Handelstage : 3 Nächster Handelstag

Abbildung 9.11: Graphische Oberfläche von Versuchsbedingung B3

Du erhältst nun zusätzlich Informationen über die **Stabilität der Indikatoren**. Die Stabilität, eine Zahl zwischen 0 und 100, gibt an, an wie vielen von 100 Handelstagen ein Indikator für den aktuellen Tag denselben Hinweis (bei demselben Fonds) liefert, wie für den vergangenen Tag. Eine Stabilität von 100 bedeutet, dass der Indikator für jeden Tag dieselbe Vorhersage macht, d.h. dieselbe Information liefert. Eine Stabilität von 50 bedeutet, dass der Indikator in der Hälfte der Tage eine andere Information liefert als am Vortag. Die Trefferquoten der vier Indikatoren bleiben dieselben.

Beispiele:

Stabilität = 50
 Stabilität = 70
 Stabilität = 100

Abbildung 9.12: Instruktion zu Bedingung B4

Informationen	Trefferrate	Stabilität	Fonds X mit S = 80	Fonds Y mit S = 80
Indikator A	80	95	<input type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
Indikator B	75	95	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
Indikator C	90	65	<input checked="" type="checkbox"/> <input type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
Indikator D	85	65	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
Anzahl vergangene Tage seit Erhalt der Information			4 3 2 1 0	4 3 2 1 0

Handel
Budget 10'000 minus Suchkosten 1'000 = investierbarer Betrag 9'000 ? → Fonds X 9'000 ? → Fonds Y 0 → Sparbüchse 0 → Auswahl bestätigen

Übersicht
Bisheriger Kontostand : 0 Gewinn in den letzten 5 Tagen : 84'000 (darunter 2 Tage ohne Gewinn) Neuer Kontostand : 84'000 Anzahl vergangener Handelstage : 4 Nächster Handelstag

Abbildung 9.13: Graphische Oberfläche von Versuchsbedingung B4

Tabelle 9.1 gibt einen Überblick über alle sechs Experimentalbedingungen und ihre wichtigsten Parameter.

Tabelle 9.1: Überblick über alle sechs Experimentalbedingungen und ihre wichtigsten Parameter

Exp.-Bed.	Zeit-achse	Alternativen-angebot	V _{Ind} (Indikatoren 1 - 4)	S _{Ziel} (X)	S _{Ziel} (Y)	S _{Ind} (Indikatoren 1–4)	Cue-Kosten	Budget
Q1	Nein	variabel	.90 ; .85 ; .80 ; .75	---	---	---	1h	240h (total)
Q2	Ja	konstant	.90 ; .85 ; .80 ; .75	---	---	---	1h	240h (total)
Q3	Ja	konstant	.90 ; .85 ; .80 ; .75	.65	.90	---	1h	240h (total)
B1	Nein	variabel	.90 ; .85 ; .80 ; .75	---	---	---	CHF 1'000	CHF 10'000
B3	Ja	konstant	.90 ; .85 ; .80 ; .75	.80	.80	---	CHF 1'000	CHF 10'000
B4	Ja	konstant	.90 ; .85 ; .80 ; .75	.80	.80	.60 ; .60 ; .95 ; .95	CHF 1'000	CHF 10'000

9.3 Experimentalablauf

9.3.1 Überblick über den Ablauf

Zu Beginn des Experiments erfassen wir den Namen der Versuchsperson sowie des Versuchsleiters / der Versuchsleiterin. Anschliessend wird ein kurzer Begrüssungstext eingeblendet (Abbildung 9.14). Danach geht es gleich los mit der ersten Versuchsbedingung.

Liebe Versuchsteilnehmerin, Lieber Versuchsteilnehmer

Herzlich willkommen bei unserem Experiment und vielen Dank für deine Bereitschaft zur Teilnahme.

Dies ist ein Entscheidungsspiel, das aus zwei Experimenten mit jeweils verschiedenen Bedingungen besteht. Es wird dir jeweils gesagt, wann das nächste Experiment oder die nächste Bedingung beginnt.

Für jedes Experiment findet später eine Verlosung unter dem erfolgreichsten Drittel aller VersuchsteilnehmerInnen statt, wobei CHF 100.- pro Experiment verlost werden.

Was dich in diesem Experiment erwartet und was du tun musst, steht auf den folgenden Seiten in der Anleitung zum Experiment.

Drücke bitte die Taste "Weiter", um zur Anleitung zu gelangen !

Abbildung 9.14: Begrüssungstext

Die 3 Bedingungen von Experiment 1B folgen direkt, d.h. ohne zeitlichen Unterbruch, auf die 3 Bedingungen von Experiment 1A. Die insgesamt sechs Bedingungen werden so, wie gerade in Kapitel 2 beschrieben, von jeder Versuchsperson in genau derselben Reihenfolge streng sequenziell absolviert: Q1, Q2, Q3, B1, B3, B4. (Es kann, mit leicht modifizierten initialen Instruktionen, auch mit den Bedingungen zum Börsenszenario begonnen werden.) Der Ablauf sämtlicher Experimentalbedingungen ist gegliedert in die drei Phasen Informationssuche, Entscheiden und ggf. Feedback. Dementsprechend ist die Bildschirmseite jeder Bedingung unterteilt in die drei Bereiche „Analyse“, „Handlung“ und „Übersicht“ (Trinkwasser-Szenario) bzw. „Informationen“, „Handel“ und „Übersicht“ (Börsenszenario).

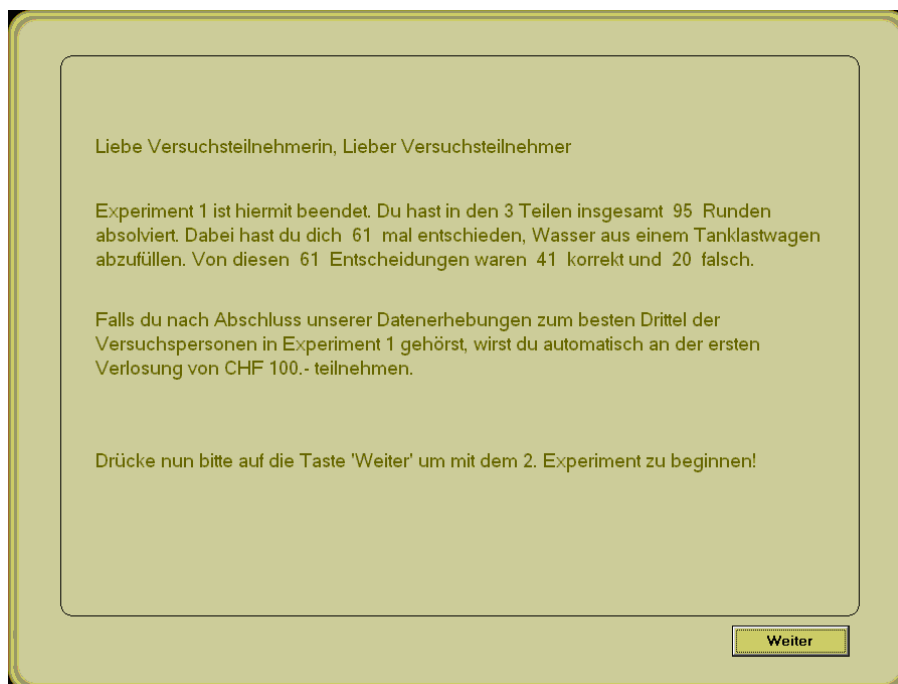


Abbildung 9.15: Resultatseite nach Q3 (exemplarische Darstellung)

Liebe Versuchsteilnehmerin, Lieber Versuchsteilnehmer

Experiment 2 ist hiermit beendet. Du hast in den 3 Teilen insgesamt 60 Runden absolviert. Dabei hast du dich 36 mal entschieden, in einen Optionsschein bzw. einen Fonds zu investieren. Von diesen 36 Entscheidungen waren 27 korrekt und 9 falsch.

Falls du nach Abschluss unserer Datenerhebungen zum besten Drittel der Versuchspersonen in Experiment 2 gehörst, wirst du automatisch an der zweiten Verlosung von CHF 100.- teilnehmen.

Weiter

Abbildung 9.16: Resultatseite nach B4 (exemplarische Darstellung)

Abschliessende Angaben

Angaben zur Person

Geburts-Datum: 08.03.1971 Geschlecht: männlich
Beruf / Studienfach + Semester: Psychologie / 3. Semester

Zusätzliche Angaben

☒ Sofern ich zum besten Drittel der Versuchspersonen gehöre, möchte ich an der Verlosung teilnehmen.
Ich kann mir vorstellen, auch später wieder einmal an einem eurer Experimente teilzunehmen ☒ Ja ☐ Nein

Email-Adresse: muster@muster.com Die Email-Adresse wird lediglich für eine allfällige Kontaktaufnahme benötigt und nicht weitergegeben.

Kommentare

Mir ist bei dem Experiment folgendes aufgefallen / Vorschläge / Anmerkungen / etc:

Weiter

Abbildung 9.17: Bildschirmseite zur Erfassung persönlicher Angaben

Im Trinkwasser-Szenario steht den Versuchspersonen pro Versuchsbedingung ein begrenzter Zeitvorrat von 240 Stunden zur Verfügung. Dadurch ist die Anzahl Runden pro Bedingung variabel. Die Dauer des Börsenszenarios ist von vornherein beschränkt auf genau 20 Runden (Handelstage) pro Bedingung. Pro Handelstag erhalten die Versuchspersonen ein virtuelles Spielgeld von CHF 10'000.-. Von diesem Betrag werden die in der aktuellen Runde anfallenden Informati-

onssuchkosten abgezogen. Der verbleibende Betrag muss innerhalb der aktuellen Runde restlos investiert werden. Er kann auf eine der beiden zur Verfügung stehenden Alternativen gesetzt oder direkt auf das Sparbuch (Gewinnkonto) transferiert werden. Weder im Trinkwasser- noch im Börsenszenario müssen die Versuchspersonen Übungsrunden absolvieren.

Nach Abschluss der jeweils letzten Versuchsbedingung eines Experiments (d.h. nach Q3 und nach B4) wird eine Seite mit den erzielten Resultaten eingeblendet. Diese umfasst die Anzahl absolvierter Runden, die totale Anzahl der Entscheidungen zugunsten der beiden Alternativen (da auch Runden absolviert werden können, in denen keine der beiden Alternativen ausgewählt wird) sowie die Anzahl korrekter und falscher Entscheidungen. Exemplarische Darstellungen der Resultatseiten sind in Abbildung 9.15 (Quellen-Szenario) und Abbildung 9.16 (Börsenszenario) zu sehen.

Nach der Präsentation des im Börsenszenario erreichten Resultats bedanken wir uns auf der nächsten Bildschirmseite für die Teilnahme an den beiden Experimenten und weisen auf die sogleich folgende Erfassung persönlicher Angaben hin. Die Erfassung dieser Daten erfolgt mittels einer einfachen Eingabemaske auf der nächsten Bildschirmseite (siehe Abbildung 9.17). Die Maske kann von den Versuchspersonen eigenständig ausgefüllt werden. Sobald dies geschehen ist, wird eine letzte Bildschirmseite angezeigt, auf der wir uns noch einmal für die Teilnahme am Experiment bedanken und die Versuchspersonen bitten, sich bei der Versuchsleitung zu melden.

9.3.2 Informationssuche (Bereich „Analyse“ bzw. „Informationen“)

Der Bereich „Analyse“ bzw. „Informationen“ dient der Informationssuche. In ihm wird das Information Board dargestellt. Im Trinkwasser-Szenario folgt noch vor dem Bereich „Handlung“, eine Legende zur Erklärung der Symbole, mit denen die Resultate der Wasserqualitätstests dargestellt werden: Ein rotes Kreuz (bestehend aus zwei diagonalen Linien) für „schlechte Wasserqualität“ und ein grüner Haken für „gute Wasserqualität“.

Über das Information Board können sich die Versuchspersonen Informationen beschaffen. Diese können zwar in freier Kombination, jedoch immer nur sequentiell, d.h. einzeln hintereinander aufgedeckt werden. Eine aufgedeckte Information bleibt bis ans Ende des momentanen Durchgangs auf dem Information Board sichtbar.

Insgesamt stehen vier verschiedene Cues zur Verfügung (in jeder Runde dieselben). Die Cues liefern eine probabilistische Aussage über die Wasserqualität einer Quelle (Trinkwasser-Szenario) bzw. über das mögliche Steigen oder Fallen eines Optionsscheins/Fonds (Börsenszenario). Sie sind im Trinkwasser-Szenario versehen mit dem Label „Test ...“ und im Börsenszenario mit dem Label „Indikator ...“. In sämtlichen Bedingungen geben wir die Validitäten der Cues als „Trefferrate“ explizit bekannt, und zwar in der zweiten Spalte des Information Boards.

In den Bedingungen Q2, Q3, B3 und B4 sehen die Versuchspersonen zusätzlich die in den letzten vier Runden gesuchten und erhaltenen Cue-Informationen pro Alternative. Mit jeder Runde wandern sämtliche Cue-Informationen vergangener Runden um ein Feld nach links, so dass sie schliesslich nach 5 Runden nicht mehr auf dem Bildschirm sichtbar sind. Zudem geben wir in

der ersten Zeile des Information Boards die Stabilitätsraten der beiden Alternativen an (jeweils rechts neben der Alternativen-Bezeichnung). In Bedingung B4 geben wir ausserdem die Stabilitätsraten der vier Cues bekannt (jeweils rechts neben der Validitätsangabe).

9.3.3 Entscheidungsphase (Bereich „Handlung“ bzw. „Handel“)

Anschliessend an die Phase der Informationsbeschaffung (= Such- und Stopp-Phase) treffen die Versuchspersonen im Bereich „Handlung“ bzw. „Handel“ ihre Entscheidung. Das Ende der Informationsbeschaffungsphase ist bestimmt durch die Auswahl einer Entscheidungsalternative. Nachdem eine der zur Verfügung stehenden Alternativen ausgewählt wurde, können keine Informationen mehr gesucht werden (die entsprechenden Schaltknöpfe sind inaktiv).

Im Trinkwasser-Szenario müssen die Versuchspersonen entscheiden, ob sie eine von zwei Tankladungen für die Herstellung von Trinkwasser verwenden wollen, im Börsenszenario, ob sie in einen von zwei Optionsscheinen (bzw. Fonds) investieren wollen. Setzen sie auf eine schlechte Alternative, d.h. auf eine Tankladung, die zur Herstellung von Trinkwasser ungeeignet ist, oder auf einen Optionsschein (bzw. Fonds), dessen Kurswert sinkt, so verlieren sie im Trinkwasser-Szenario 10 Stunden Zeit (da sie für ihren Fehlentscheid keinen Punkt erhalten) und im Börsenszenario den investierten Betrag. Setzen sie hingegen auf eine gute Alternative (also eine brauchbare Tankladung bzw. einen Optionsschein/Fonds, der im Wert steigt), so erhalten sie im Trinkwasserszenario einen Gewinnpunkt (in dem Sinne, dass der Zählerstand der guten Entscheidungen um einen Punkt ansteigt) und im Börsenszenario wird ihnen das Dreifache des eingesetzten Betrags auf ihr Gewinnkonto gutgeschrieben. In allen Bedingungen beginnen die Versuchspersonen jeweils in Runde 1 mit einem Punkte- bzw. Gewinnkontostand von null Punkten bzw. Franken.

Falls eine Versuchsperson keine der beiden angebotenen Alternativen für aussichtsreich hält, steht ihr noch eine dritte Möglichkeit offen: Sie kann beide Tankladungen wegschütten (Trinkwasser-Szenario) bzw. den investierbaren Betrag direkt auf das Gewinnkonto transferieren (Börsenszenario). Dadurch geht keine Zeit verloren bzw. wächst der Gewinnkontostand immerhin noch um den Betrag des investierbaren Spielgelds (d.h. das Runden-Spielgeld abzüglich Suchkosten) an.

Den Versuchspersonen wird kein expliziter Zielzustand vorgegeben. Sie werden lediglich angewiesen, jeweils möglichst viele richtige Entscheidungen zu treffen, d.h. möglichst viele Punkte bzw. einen möglichst hohen Spielgewinn zu erreichen.

Eine einmal getroffene Entscheidung kann beliebig oft geändert werden. Erst nach Drücken der Schaltfläche „Auswahl bestätigen“ kann die getroffene Entscheidung definitiv nicht mehr rückgängig gemacht werden. Die Schaltfläche wird daraufhin automatisch deaktiviert. In der Folge werden nun ggf. die im Übersichtsteil angezeigten Informationen (Spielstand, etc.) auf den neuesten Stand gebracht.

9.3.4 Feedback (Bereich „Übersicht“)

Nach jeweils drei bis fünf Runden – das Intervall wird jedes Mal neu durch Zufall bestimmt – sowie in der letzten Runde einer jeden Bedingung erhalten die Versuchspersonen Feedback über die im letzten Intervall erzielte Leistung: Im Trinkwasser-Szenario informieren wir die Versuchspersonen über die seit dem letzten Feedback neu hinzugekommene Anzahl brauchbarer Wasserlieferungen sowie die Anzahl brauchbarer Wasserlieferungen insgesamt. Im Börsenszenario teilen wir den Versuchspersonen die Höhe des im Intervall erzielten Gewinns sowie des bisher erreichten Gesamtgewinns mit. Ausserdem informieren wir über die Dauer des abgelaufenen Intervalls und die Anzahl Tage (im Intervall) ohne Gewinn.

Die im Trinkwasser-Szenario angegebenen Informationen über den aktuellen Zeitverbrauch und die noch vorhandene Zeit repräsentieren immer den aktuellen Zustand. Diese Informationsfelder werden fortlaufend, d.h. nach jedem Cuekauf bzw. jeder definitiven Entscheidung aktualisiert. Dasselbe gilt für den Zähler „Anzahl vergangener Handelstage“ im Börsenszenario.

9.4 Technische Spezifikation und Datenbasis

9.4.1 Technische Spezifikation

Cue-Bezeichnungen. Die Cue-Bezeichnungen stehen bei jedem Experiment und bei allen Bedingungen fix an derselben Stelle auf dem Bildschirm (d.h. "Test A" bzw. „Indikator A“ steht immer an oberster Stelle, "Test B" bzw. „Indikator B“ immer an zweitoberster Stelle, etc.). Die hinter einer bestimmten Cue-Bezeichnung liegenden Cue-Daten sind jedoch in ihrer Reihenfolge von Proband zu Proband randomisiert (aber innerhalb einer Versuchsperson für alle drei Bedingungen eines Szenarios fix, um die Versuchsperson nicht unnötig zu verwirren).

Cue-Validitäten: Der Bereich möglicher Cue-Validitäten erstreckt sich von .50 bis 1.0. Eine Cue-Validität von .50 bedeutet, dass die vom Cue gelieferte Information mit jeweils gleich hoher Wahrscheinlichkeit korrekt oder falsch ist und man anstelle der Konsultation eines solchen Cues genauso gut raten könnte. Die Validitäten der vier Cues sind über alle Bedingungen und Versuchspersonen konstant gehalten, um die Vergleichbarkeit der Versuchsbedingungen gewährleisten zu können. Sie sind in 5%-Schritten abgestuft und im Hinblick auf Versuchsbedingungen mit Zeitverlauf relativ hoch gewählt (.90, .85, .80, .75). Dadurch stellen wir sicher, dass eine Cue-Information, bei gegebener Zielvariablen-Stabilität von .95, auch nach zwei Zeitschritten (Runden) noch eine Validität von mindestens .67 aufweist und damit für die Versuchspersonen potentiell noch immer attraktiv ist (siehe Tabelle 9.2).

Die angegebenen Validitäten entsprechen exakt der tatsächlichen Wahrscheinlichkeit, mit der sie die Ausprägungen der Zielvariablen korrekt vorhersagen. Die Berechnung erfolgt immer über genau 20 Runden und jeweils für beide zur Auswahl stehenden Alternativen getrennt. Wir haben darauf geachtet, dass die getrennt pro Alternative berechnete Cue-Validität genau der über beide Alternativen gemittelten Validität des Cues entspricht. Für die Versuchsperson sind diese Daten also absolut verlässlich.

Tabelle 9.2: Cue-Validitäten in Abhängigkeit der Zielvariablen-Stabilität (nach einem bis vier Zeitschritten)

Zielvariablen-Stabilität = .65					
Cue	V	1	2	3	4
1	.90	.585	.491	.462	.454
2	.85	.553	.463	.436	.428
3	.80	.520	.436	.411	.403
4	.75	.488	.409	.385	.378
Zielvariablen-Stabilität = .80					
Cue	V	1	2	3	4
1	.90	.720	.612	.547	.508
2	.85	.680	.578	.517	.480
3	.80	.640	.544	.486	.452
4	.75	.600	.510	.456	.424
Zielvariablen-Stabilität = .95					
Cue	V	1	2	3	4
1	.90	.855	.815	.778	.745
2	.85	.808	.769	.735	.704
3	.80	.760	.724	.692	.662
4	.75	.713	.679	.648	.621

Cue-Stabilitäten. Im Trinkwasser-Szenario beträgt die Stabilität der beiden valideren Cues (mit $V=.90$ bzw. $V=.85$) .65, die Stabilität der beiden weniger validen Cues (mit $V=.80$ bzw. $V=.75$) .90. Im Börsenszenario hingegen beträgt die Stabilität der beiden weniger validen Cues .95. Bei der Stabilität von .65 handelt es sich um einen Mittelwert über beide Zielvariablen. Pro Zielvariable gemessen betragen die Stabilitäten der entsprechenden Cues entweder .60 oder .70.

Die Stabilitäten der Cues sind gegenläufig zu ihren Validitäten. Dank dieser Gegenläufigkeit lässt sich klar unterscheiden, ob eine Versuchsperson eher valide oder eher stabile Cues präferiert. Die Cue-Stabilität von .90 bzw .95 ist bewusst hoch gewählt: Bei einer Cue-Stabilität von .95 liegt die Validität einer zwei Tage alten Information - abhängig von der Validität der aktuellen Information - noch immer zwischen .68 und .82.

Zielvariablen-Stabilitäten. Die Stabilitäten der Zielvariablen betragen in allen drei Bedingungen des Trinkwasser-Szenarios .60 für Alternative X und .90 für Alternative Y. Im Börsenszenario hingegen in allen drei Bedingungen .80 für beide Alternativen.

Anzahl Versuchsdurchgänge pro Bedingung. Um für statistische Auswertungen pro Versuchsperson eine ausreichende Menge an Daten zu erhalten, absolviert jede Versuchsperson pro Versuchsbedingung im Minimum 20 Durchgänge.

Im Trinkwasser-Szenario stehen jeweils genau 240 Stunden Zeit als Ressourcenvorrat pro Versuchsbedingung zur Verfügung. Pro Runde können maximal 18 Stunden Spielzeit „verbraucht“ werden: 8 x 1 Stunde für die Informationssuche und 10 Stunden für das Abfüllen einer Tankkla-

dung. Dadurch erreichen wir, dass sich eine Versuchsbedingung über mindestens 13 Runden erstreckt ($240\text{h} / 18\text{h} = 13.3$). Tatsächlich aber erwarten wir eine höhere Anzahl an minimal absolvierten Runden, da wir davon ausgehen, dass die Versuchspersonen nie oder nur in seltenen Fällen alle 8 Cues pro Durchgang konsultieren. Wir gehen also davon aus, dass trotz der Möglichkeit, weniger als 20 Runden pro Bedingung zu absolvieren, jede Versuchsperson mindestens 20 Runden durchläuft.⁴⁵

Im Börsenszenario ist die Anzahl Durchgänge von vornherein auf genau 20 Runden pro Bedingung festgelegt.

Im Trinkwasser-Szenario können Versuchsdurchgänge quasi „übersprungen“ werden, indem weder nach Cue-Informationen gesucht noch Wasser abgefüllt wird und folglich null Stunden Zeit vom Gesamtzeitvorrat abgezogen werden. (In den Bedingungen Q2 und Q3 kann das durchaus eine sinnvolle Strategie sein, wenn man nämlich zuvor negative Testresultate erhielt und darauf spekuliert, dass sich die Wasserqualität erst einige Tankladungen später ändern wird.) Im Börsenszenario hingegen muss in jeder Runde ein konstanter Ressourcenbetrag von CHF 10'000.- vollständig investieren werden.

Ausprägungsverläufe der Zielvariablen: Basisraten. Die Basisrate der positiven Ausprägungen der Zielvariablen ist in allen sechs Versuchsbedingungen identisch und beträgt jeweils $1/3$. Das heisst: 6-7 von 20 Tankladungen (einer jeden Alternative) enthalten für die Trinkwasserproduktion geeignetes Wasser bzw. in 6-7 von 20 Handelstagen (pro Alternative) steigt der Wert des Optionsscheins/Fonds. Damit ist sichergestellt, dass durch eine Ratestrategie grundsätzlich keine gute Performance erreicht werden kann.

Punktezahl pro Versuchsdurchgang. Im Trinkwasser-Szenario könnten pro Versuchsbedingung in den 240 Stunden theoretisch maximal 24 Punkte erreicht werden. Praktisch sind in allen drei Trinkwasser-Bedingungen jeweils lediglich maximal 10 Punkte pro gespielten 20 Runden möglich, und zwar aus den folgenden beiden Gründen: Erstens ist die Basisrate der positiven Ausprägungen für die beiden Zielvariablen auf je ein Drittel festgelegt (siehe oben). Das bedeutet, dass lediglich 6 bis 7 von 20 Tankladungen (pro Alternative) Wasser von brauchbarer Qualität enthalten. Zweitens überlagern sich in jeder Versuchsbedingung die positiven Ausprägungen der beiden Zielvariablen im zeitlichen Verlauf. Diese Überlagerung ist à priori so festgelegt, dass in 6 bis 7 der 20 Runden genau einer der beiden Tanks, Wasser von ausreichender Qualität enthält, in 3 bis 4 Runden beide und in genau 10 Runden jeweils keiner von beiden.

Allerdings ist weder die Anzahl der Runden auf 20 beschränkt, noch ist es zwingend nötig, sich in jeder Runde für eine der beiden Alternativen zu entscheiden. Mit dem Wegkippen beider Tankladungen zum richtigen Zeitpunkt kann so die Trefferzahl erhöht werden. Im Börsenszenario ist das nicht der Fall. Hier lässt sich jedoch das Gewinnkonto maximieren, indem zur richti-

⁴⁵ Diese Erwartung bestätigte sich, indem in der ersten Durchführung des Experiments nur eine einzige Versuchsperson in einer der sechs Bedingungen weniger als 20 Runden absolviert hatte (nämlich nur 17).

gen Zeit Geld direkt transferiert wird, ohne auf Optionsscheine/Fonds zu setzen. Die Zahl möglicher Optionsschein-/Fonds-Gewinne ist hier jedoch klar auf 10 begrenzt.

Zeitverlauf und Gedächtnisbelastung: Dank der Zeitverlaufsdarstellung der gesuchten (und erhaltenen) Cue-Informationen brauchen die Versuchspersonen diese Informationen nicht im Gedächtnis zu behalten. In den Versuchsbedingungen mit einer Zeitverlaufsdarstellung wird jeweils der Verlauf der Informationssuche des aktuellen Durchgangs sowie der letzten vier Runden angezeigt. Die Beschränkung auf die Präsentation von Informationen, die maximal vier Runden alt sind, besteht, weil die Validitätswerte von noch älteren Informationen aufgrund der Stabilitäten der Cues äusserst gering ausfallen (So gering, dass sie nicht mehr in die Überlegungen für die aktuelle Informationssuche miteinbezogen werden sollten). Bei entsprechend guter Gedächtnisleistung der Versuchspersonen wäre dies aber trotzdem möglich, wenn auch nicht sinnvoll.

Instruktionen. Vor jeder Bedingung werden die Versuchspersonen mittels einer auf dem Bildschirm präsentierten Anleitung instruiert, um was es geht und was sie tun sollen. Die Instruktionen sind pro Bedingung in eine kurze Geschichte integriert. Diese dient jeweils dazu, die Versuchspersonen auf eine möglichst plausible, zwar erfundene aber dennoch realitätsnah ausgestaltete Arbeitssituation einzustimmen. Sämtliche Instruktionstexte sind bei der Beschreibung des Experiments in Kapitel 2 und 3 bereits im Wortlaut aufgeführt. Die Geschichten so ausgestaltet, dass nach deren Lektüre möglichst keine Fragen bezüglich ihrer Plausibilität auftauchen.

Zeitbedarfsabschätzung. Das gesamte Experiment soll eine Länge von 60 Minuten nicht überschreiten. Die Abschätzung des realen Zeitbedarfs pro Versuchsbedingung ist demnach wichtig für die Bestimmung der Anzahl insgesamt durchführbarer Bedingungen pro Versuchsperson. Die Dauer einer Bedingung hängt ihrerseits massgeblich von der Anzahl absolvierter Runden in dieser Bedingung ab.

Für das Börsenszenario fällt diese Abschätzung etwas leichter, da hier die Anzahl Runden pro Bedingung konstant ist: Bei einer durchschnittlichen Zeit von 15 Sekunden pro Runde benötigt eine Versuchsperson für alle 20 Runden einer Bedingung im Mittel ca. 5 Minuten. Inklusive 1 bis 2 Minuten für das Durchlesen und Verstehen der Instruktion ergibt dies ca. 6 bis 7 Minuten pro Bedingung bzw. total 18 bis 21 Minuten für alle 3 Bedingungen zusammen.

Für das Trinkwasser-Szenario gehen wir von bis zu 35 absolvierten Runden aus. Bei 2 Minuten Instruktionszeit pro Bedingung und durchschnittlich 15 Sekunden Realzeitverbrauch pro Runde ergibt dies ca. 11 Minuten pro Bedingung bzw. total 33 Minuten für alle 3 Bedingungen zusammen. Inklusive Begrüssung, Erhebung von Angaben zur Person und Verabschiedung dürfte der Zeitbedarf für das gesamte Experiment die Vorgabe von 60 Minuten also nicht überschreiten.

Tabelle 9.3: Datenbasis (Ausprägungsverläufe der Zielvariablen und Cues für alle sechs Versuchsbedingungen. 1 = negative Ausprägung; 2 = positive Ausprägung)

Q1 Alternative X																				
t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Zielvar.	1	1	1	1	1	1	2	1	1	2	1	2	2	2	1	2	2	1	1	1
Wechsel	0	0	0	0	0	0	1	1	0	1	1	1	0	0	1	1	0	1	0	0
Cue1	1	1	1	1	1	1	2	1	1	2	1	2	2	2	2	2	2	1	2	1
Cue2	1	1	2	1	1	1	1	1	1	2	2	2	2	2	1	2	2	1	1	1
Cue3	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	1	1
Cue4	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	1	1

Alternative Y																				
t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Zielvar.	1	1	1	1	1	1	1	2	2	2	2	2	2	2	1	1	1	1	1	1
Wechsel	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0
Cue1	1	1	1	2	1	1	1	2	2	2	2	2	2	2	1	2	1	1	1	1
Cue2	1	1	1	1	1	1	1	2	1	2	1	2	2	2	1	1	1	2	1	1
Cue3	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	1	1	1
Cue4	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	1	1	1

Q2 Alternative X																				
t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Zielvar.	1	1	1	2	2	1	2	2	2	1	2	1	1	2	1	1	1	1	1	1
Wechsel	0	0	1	0	1	1	0	0	1	1	1	0	1	1	0	0	0	0	0	0
Cue1	1	2	1	2	2	2	2	2	2	1	2	1	1	2	1	1	1	1	1	1
Cue2	1	1	1	2	2	1	2	2	2	2	2	1	1	1	1	1	2	1	1	1
Cue3	1	1	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1
Cue4	1	1	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1

Alternative Y																				
t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Zielvar.	1	1	1	1	1	1	1	2	2	2	2	2	2	2	1	1	1	1	1	1
Wechsel	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Cue1	1	1	1	2	1	1	2	2	2	2	2	2	2	2	1	1	1	2	1	1
Cue2	1	1	2	1	1	1	2	2	2	1	2	1	2	1	1	1	1	1	1	1
Cue3	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	1	1
Cue4	1	1	1	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1

Q3 Alternative X																				
t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Zielvar.	1	1	1	1	1	1	2	1	1	2	1	2	2	2	1	2	2	1	1	1
Wechsel	0	0	0	0	0	1	1	0	1	1	1	0	0	1	1	0	1	0	0	0
Cue1	1	1	1	1	1	1	2	1	1	2	1	2	2	2	2	2	2	1	2	1
Cue2	1	1	2	1	1	1	1	1	1	2	2	2	2	2	1	2	2	1	1	1
Cue3	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	1	1
Cue4	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	1	1

Alternative Y																				
t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Zielvar.	1	1	1	1	1	1	1	2	2	2	2	2	2	2	1	1	1	1	1	1
Wechsel	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Cue1	1	1	1	2	2	2	2	1	1	1	2	1	1	2	2	2	2	2	1	1
Cue2	1	1	1	2	2	2	1	1	1	2	1	1	1	2	2	1	2	2	1	1
Cue3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2
Cue4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2

B1 Alternative X																				
t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Zielvar.	1	1	2	2	2	2	2	2	1	1	1	1	1	1	1	2	1	1	1	1
Wechsel	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	1	1	0	0	0
Cue1	1	1	2	2	2	2	1	2	1	1	1	1	1	1	1	2	1	1	2	1
Cue2	1	2	2	2	2	2	2	1	1	1	2	1	1	2	2	1	1	1	1	1
Cue3	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1	1	1
Cue4	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Alternative Y																				
t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Zielvar.	1	1	1	1	2	1	1	1	1	1	1	1	1	2	2	2	2	2	1	1
Wechsel	0	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0
Cue1	1	2	1	1	2	1	1	1	1	1	1	1	2	1	2	2	2	2	2	1
Cue2	1	1	1	2	2	1	1	2	1	1	1	2	1	1	2	2	1	2	2	1
Cue3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2
Cue4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2

B3 Alternative X																				
t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Zielvar.	1	1	2	2	2	2	2	1	1	1	1	1	1	1	2	2	1	1	1	1
Wechsel	0	1	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0
Cue1	1	1	2	2	2	2	2	1	1	2	1	1	1	2	2	2	1	1	1	1
Cue2	1	1	2	2	1	2	2	1	1	1	2	1	1	1	2	2	2	1	1	1
Cue3	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1
Cue4	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1

Alternative Y																				
t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Zielvar.	1	1	1	1	2	1	1	1	1	1	1	1	1	2	2	2	2	2	1	1
Wechsel	0	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0
Cue1	1	2	1	1	2	1	1	1	1	1	1	1	2	1	2	2	2	2	2	1
Cue2	1	1	1	2	2	1	1	2	1	1	1	2	1	1	2	2	2	2	2	2
Cue3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2
Cue4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2

B4 Alternative X																				
t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Zielvar.	1	1	2	2	2	2	2	2	1	1	1	1	1	1	2	1	1	1	1	1
Wechsel	0	1	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0
Cue1	1	1	2	2	2	2	1	2	1	1	1	1	1	1	2	1	1	2	1	1
Cue2	1	2	2	2	2	2	2	1	1	1	2	1	1	2	2	1	1	1	1	1
Cue3</																				

verlauf einer Zielvariablen ein 100% stabiler Cue unmöglich eine Validität von 1.0 haben kann. Setzen wir natürlicherweise den Ausprägungsverlauf einer Zielvariablen (und damit auch deren mittlere Stabilität) als gegeben voraus, so wird der Bereich möglicher Stabilitätswerte, die ein Cue noch annehmen kann, determiniert durch seine Validität und umgekehrt: Der Bereich möglicher Validitäten eines Cues ist determiniert durch seine eigene Stabilität und die Stabilität der Zielvariablen, deren Ausprägungen er vorhersagen soll.

Festlegung der Ausprägungsverläufe der Zielvariablen und der Cues. Für alle drei Quellenbedingungen gilt, dass die Validitäten der Cues keine Extremwerte annehmen. Dasselbe gilt (in Q3) auch für die Stabilitäten der Zielvariablen. Dadurch erhalten wir grössere Freiheiten in der Ausgestaltung der Ausprägungsverläufe der Cues bzw. der Zielvariablen.

Die Datenbasis der sechs Bedingungen (Ausprägungsverläufe der Zielvariablen und Cues) ist in Tabelle 9.3 dokumentiert. Für jede Versuchsbedingung haben wir eine Serie von 20 Durchgängen erstellt. Im Trinkwasser-Szenario, in dem eine Versuchsbedingung mehr als 20 Runden haben kann, werden die Daten ggf. rezykliert (d.h. für die 21. Runde verwenden wir wieder die Daten aus dem ersten Datensatz etc.). Somit kann eine Versuchsbedingung beliebig viele Durchgänge haben, ohne dass dabei die Daten ausgehen.

9.6 Schlussbemerkungen

Das hier vorgestellte Experimentaldesign ermöglicht es, den Faktor Zeitverlauf in Untersuchungen zur Informationssuche beim Entscheiden einfließen zu lassen, ohne den vertrauten Weg des information board dabei zu verlassen. Es handelt sich unseres Wissens um das erste Design, das in diese Richtung ausgelegt ist. Zwingend einzubeziehende Faktoren sind in einem solchen Ansatz die zeitliche Stabilität der Zielvariablen (Q2, Q3, B3, B4) und – um Informationen mit Blick auf die Zukunft suchen zu können – auch die der Cues (B4). Dabei ist a priori nicht vorherzusagen, ob Personen überhaupt die Stabilität eines Cues als eine der Validität vergleichbare Cue-Qualität wahrnehmen (Vergleich B3 / B4) und ob sie auf Unterschiede in der Stabilität überhaupt reagieren (Vergleich Q3 / B3). Daneben wird Zeit natürlich in einem solchen Kontext selbst zu einer Ressource, die nicht unbedingt dieselben Qualitäten hat wie die in der Entscheidungspsychologie übliche Ressource Geld besitzt (oder neutral Punkte, welche aber beliebig in Geldbeträge übertragbar wären). Aus diesem Grund ist auch diese Ressourcenvariable variiert worden, was mit dem Trinkwasser-Szenario eine ganz neue Experimentalidee erforderte (das Börsenszenario leitet sich ja unmittelbar aus bisherigen Designs ab und erweitert diese lediglich).

All diese Variationen waren so in ein Gesamtexperiment zu verpacken, dass die Einzelwirkung der Faktoren zumindest explorativ überprüfbar ist. Ausserdem sollten, soweit in der Begrenzung auf ein Experiment der praktikablen Länge von 60 Minuten überhaupt möglich, auch Kombinationen der Faktoren überprüfbar und die einzelnen Bedingungen untereinander vergleichbar sein. Und um überhaupt eine Wirkung eines der mit dem Zeitverlauf neu eingeführten Faktoren überprüfen zu können, musste natürlich auch für jedes der beiden Szenarien eine „baseline“ (ohne die

sich ergebenden Abhängigkeiten zeitlicher Verläufe, aber in sich gleichwohl für die Probanden logisch) in das Gesamtdesign eingebaut werden.

All dies unterliegt in der Summe natürlichen Limitationen, die für ein derartiges Experiment in Kauf genommen werden müssen. So ist beispielsweise die Vergleichbarkeit zwischen Q3 und B3 nicht perfekt gegeben, da sich die beiden Varianten neben der unterschiedlichen Ressource (Zeit oder Geld) auch in den Stabilitäten der Zielvariablen (beide gleich bzw. verschieden) unterscheiden. Die Forschungsberichte, die sich mit den einzelnen Fragestellungen und deren Beantwortungen beschäftigen, listen denn auch eine Reihe solcher Limitationen auf (jeweils bezogen auf die spezifische Fragestellung). Gleichwohl sind wir überzeugt, mit der grundsätzlichen Entwicklung dieses Experimentalansatzes die Tür aufgestossen zu haben, um der Bedeutung des Faktors Zeit in der Entscheidungsforschung künftig mehr Rechnung verleihen zu können.

9.7 Literatur

- Christen, S., Hausmann, D. & Läge, D. (2007). *Ein Optionsscheinszenario zur Untersuchung der Strategien sequentieller Informationssuche: Ablauf, Aufbau und technische Spezifikation der Experimentalumgebung*. AKZ-Forschungsbericht 42. Zürich: Angewandte Kognitionspsychologie.
- Gigerenzer, G., Todd, P. M., et al., Eds. (1999). *Simple Heuristics That Make Us Smart*. New York: Oxford University Press.
- Maule, A. J., & Edland, A. C. (1997). *The effects of time pressure on human judgement and decision making. Decision making: cognitive models and explanations*. R. Ranyard, W. R. Crozier and O. Svenson. London: Routledge.
- Payne, J. W. (1976). Task complexity and contingent processing in decision making: An information search and protocol analysis. *Organizational Behavior and Human Performance*. 16, 366-387.
- Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. Cambridge: Cambridge University Press.
- Zurbriggen, S., Christen, S., Hausmann, D. & Läge, D. (2007a). *Informationssuche beim Entscheiden in dynamischen Umwelten*. AKZ-Forschungsbericht 44. Zürich: Angewandte Kognitionspsychologie.
- Zurbriggen, S., Christen, S., Hausmann, D. & Läge, D. (2007b). *Adaptivität der Informationssuche in statischen und in dynamischen Umwelten*. AKZ-Forschungsbericht 46. Zürich: Angewandte Kognitionspsychologie.
- Zurbriggen, S., Hausmann, D., Christen, S., & Läge, D. (2007). *Die zeitliche Dynamik von Umwelten: Heuristisches Wahlverhalten am Beispiel eines Trinkwasser-Szenarios mit und ohne Zeitachse*. AKZ-Forschungsbericht 41. Zürich: Angewandte Kognitionspsychologie.
- Zurbriggen, S. & Läge, D. (2007). *Ressource Zeit statt Geld: Ein Vergleich zweier Szenarien zur Entscheidungsforschung in dynamischen Umwelten*. AKZ-Forschungsbericht 45. Zürich: Angewandte Kognitionspsychologie.

Kapitel 10

Die zeitliche Dynamik von Umwelten: Heuristisches Wahlverhalten am Beispiel eines Trinkwasser-Szenarios mit und ohne Zeitachse

Entscheiden spielt sich im Alltag in der Regel im Rahmen von Zeitverläufen ab. Möglicherweise relevante Information stammt dabei oft aus der Vergangenheit, und es muss eingeschätzt werden, ob sich eine Aktualisierung lohnt. Eine erneute Informationssuche kann dabei im Konflikt mit anderen Aktivitäten stehen, so dass die Ressource Zeit eine Rolle für die Frage der Informationssuche beim Entscheiden spielt. Um diese alltägliche Situation für die Entscheidungspsychologie erforschbar zu machen, wurde ein Trinkwasserszenario entwickelt, welches eine entsprechende zeitlich dynamische Umwelt darstellt. Hier spielt neben der Ressource Zeit vor allem auch das Wissen um die zeitliche Stabilität der wählbaren Alternativen eine zentrale Rolle. In einem ersten Experiment mit diesem Szenario wurde nun überprüft, wie weit die Versuchspersonen auf eine solche zeitlich dynamische Umwelt reagieren (im direkten Vergleich zu einer statischen Umwelt mit Einzelentscheidungen). Die Befunde bestätigen im Wesentlichen die aufgestellten Hypothesen hinsichtlich Suchregel, Stoppregel und Entscheidungsregel der verwendbaren schrittweisen Entscheidungsheuristiken. Sie zeigen aber auch, dass zumindest ein Teil der Probanden nicht oder nur sehr wenig auf den Wechsel der statistischen zur zeitlich dynamischen Umwelt reagiert.

10.1 Einleitung

Das Leben spielt sich in einem Zeitverlauf ab, und so auch das Entscheiden. Nicht nur, dass es gute und weniger gute Zeitpunkte für eine Entscheidung gibt (vgl. Kerstholt & Raaijmakers, 1997), sondern auch die Informationen, welche einer Entscheidung zugrunde liegen, ändern sich über die Zeit. Deswegen sollte die Suche nach den probabilistischen Hinweisen auf die beste wählbare Alternative auch möglichst eng an den Zeitpunkt der Entscheidung selbst gekoppelt sein, wenn man eine maximale Validität von Prädiktorvariablen (Cues) für eine Zielvariable (Entscheidungsalternative) erreichen will.

Die bisherige experimentelle Entscheidungsforschung hat dies, soweit sie sich überhaupt mit Informationssuche als einem aktiven Prozess zur Vorbereitung einer Entscheidung beschäftigt hat, dadurch realisiert, dass Informationssuche und Entscheidung auf denselben Zeitpunkt gelegt wurden (Bröder, 2000a, 2000b, 2002, 2003, 2005; Bröder & Schiffer, 2003, 2006; Gigerenzer, Todd, & ABC Research Group, 1999; Hausmann & Läge, 2005; Läge, Hausmann, & Christen, 2005; Läge, Hausmann, Christen, & Daub, 2005; Newell, Rakow, Weston, & Shanks, 2004; Newell & Shanks, 2003; Newell, Weston, & Shanks, 2003; Payne, 1976; Payne & Bettman, 2000; Payne, Bettman, & Johnson, 1988; Payne, Bettman, & Johnson, 1993; Rieskamp & Hoffrage, 1999). Die dabei eingesetzten Information Boards waren dann entsprechend so strukturiert, dass es sich wiederholende Durchgänge gab, in denen jeweils Informationen gesucht und auf der Basis dieser Informationen entschieden werden konnte. Der nächste Durchgang hatte dann (von der Information her gesehen) nichts mehr mit dem vorausgehenden Durchgang zu tun.

Dies ist natürlich eine Abstraktion der Wirklichkeit, eine sinnvolle freilich, wenn man – wie hier im Kontext der Erforschung schrittweiser Entscheidungsheuristiken – etwas über die Reihenfolge der Informationssuche, über das Beenden dieser Suche beim Vorliegen bestimmter Informationen sowie über die Entscheidungsregel bei den dann vorliegenden Informationen wissen möchte. Die drei Schritte gelten als die Bausteine einfacher Entscheidungsheuristiken (vgl. z.B. Gigerenzer et al., 1999) und erhielten deswegen eine gewisse Aufmerksamkeit in der bisherigen experimentellen Forschung.

Auf der Basis dieser Ansätze lässt sich nun die experimentelle Wirklichkeit ein wenig näher an die Lebenswirklichkeit heranrücken, wenn man berücksichtigt, dass der Zeitpunkt der Informationssuche nicht immer der Zeitpunkt der Entscheidung ist. Häufig verfügen wir ja bereits aus der Vergangenheit über Informationen, welche wir in eine aktuelle Entscheidung mit einfließen lassen können. Diese Informationen sind allerdings, sofern sie probabilistischen Charakter besitzen, nicht mehr ganz so valide wie eine aktuell gesuchte Information zu einer oder mehreren Entscheidungsalternativen (vorausgesetzt einmal, Prädiktorvariable und Zielvariable sind nicht hundertprozentig über die Zeit stabil). Also werden wir – sofern wir mit der eingeschätzten verbleibenden Validität der Informationen aus der Vergangenheit nicht mehr zufrieden sind – unser Wissen durch eine aktuelle Informationssuche ergänzen. Dies ist ein ganz normales Verhalten in den zeitlich dynamischen Umwelten unseres Alltags und lässt sich durchaus in der experimentellen Entscheidungsforschung modellieren. Man muss nur eine Zeitachse in die beste-

henden Information Boards einbauen. Hinzutretende Untersuchungsvariablen wären dann beispielsweise der Aktualitätsgrad der Information oder – für die Abschätzung der Situation in einer zeitlich dynamischen Umwelt enorm wichtig – der Stabilitätsgrad einer Alternative.

Ausserdem wird in einer Umwelt mit zeitlicher Verlaufsperspektive auch die Zeit selbst zu einer wichtigen Ressource. Auch wenn wir uns gerne Zeit als infiniten Strahl vorstellen, so ist sie für Organismen doch grundsätzlich begrenzt. Auch für die Erledigung bestimmter Aufgaben gibt es immer spezifische zeitliche Begrenzungen. Hat man mehrere Dinge zu tun, so können diese miteinander um die durch die grundsätzliche Begrenzung entstandene Ressource Zeit konkurrieren. Das ist bereits im Entscheidungsprozess der Fall: Zeit, die jemand für die Informationssuche aufwendet, kann er nicht mehr für einen anderen Bereich einsetzen. Deswegen wird er bestrebt sein, den Prozess der Informationssuche möglichst effizient zu gestalten.

Bislang wurde in der Entscheidungsforschung dieser grundsätzlich bestehende Aufwand bei der Informationssuche durch monetäre Kosten experimentell implementiert. Das ist sinnvoll, um Aufwand und Erreichtes (der Wert von Alternativen, für die sich eine Versuchsperson entschieden hat) möglichst direkt miteinander in Beziehung setzen zu können. Grundsätzlich ist aber auch denkbar, dass man ein Experiment gestaltet, in dem die Zeit selbst als Ressource für die Informationssuche herangezogen wird. Eine solche Idee ist beispielsweise in einem Experiment realisiert, in dem ein Information Board möglichst schnell zu bearbeiten ist (z.B. Rieskamp & Hoffrage, 1999, 2006; Kerstholt, 1994). Aber natürlich braucht man hier nicht unbedingt ein „Echtzeitszenario“ einzuführen.

Im Folgenden werden wir ein Experiment vorstellen, welches eine Zeitachse implementiert und dabei auch bei der Ressource ganz auf die Zeit setzt. Dieses Experiment wird als zweite experimentelle Neuerung einen Stabilitätskoeffizienten für die möglichen Entscheidungsalternativen einführen und so den Versuchspersonen die Möglichkeit geben, die verbleibende Validität zeitlich zurückliegender Informationen korrekt einzuschätzen. Das Experiment versteht sich als Erweiterung der bisherigen Untersuchungen zur Erforschung schrittweiser Urteilsheuristiken. Aus diesem Grund wollen wir – bevor wir das Experimentaldesign selbst vorstellen – noch kurz überlegen, welche Auswirkungen die Einführung einer Zeitachse auf die drei wichtigsten Bausteine dieser Heuristiken (Suchregel, Stoppregel, Entscheidungsregel) haben.

Suchregel. Der wichtigste Index für die Qualität eines probabilistischen Cues ist seine Validität. Nicht zuletzt deswegen schlagen beispielsweise auch Gigerenzer & Goldstein (1996) vor, in einer einfachen Entscheidungsheuristik die Cues nach absteigender Validität zu befragen, bis eine diskriminierende Information gefunden wurde. Sofern man über Cues verfügt, welche unmittelbar eine Information über alle Entscheidungsalternativen geben, können diese Cues in einem grösseren Set von Alternativen nicht immer zwischen allen Alternativen diskriminieren. Vor allem wenn die Suche nach Information mit Aufwand verbunden ist, kann es ratsam sein, neben der Validität auch die Diskriminationsrate in die Überlegungen nach der optimalen Reihenfolge der zu suchenden Information einzubeziehen. Dies ist beispielsweise im Success-Koeffizienten (Newell et al., 2004) und im Usefulness-Koeffizienten (Hoffrage, Hertwig, & Czienskowski,

2003) der Fall. In beiden führt eine Verrechnung von Validität und Diskriminationsrate zum besten Index für die Festlegung einer Suchregel bei Informationskosten. Diese Koeffizienten sind aber nur in Fällen relevant, in denen die Cues nicht immer diskriminieren. Ansonsten gilt weiterhin die Validität als beste Suchregel.

Im Kontext der Einführung einer Zeitachse wird diese Validität nun massgeblich durch die Stabilität der Alternativen beeinflusst, zumindest was das Heranziehen von Cue-Informationen aus der Vergangenheit anbelangt: Bei einer über die Zeit perfekt stabilen kann man sich genauso gut auf einen gleich validen Indikator aus der Vergangenheit verlassen wie auf einen aus der Gegenwart. Je geringer die Stabilität jedoch wird, umso geringer wird auch die Vorhersagekraft vergangener probabilistischer Cue-Information für die aktuelle Ausprägung des Cues. Und je mehr Zeit (hier Anzahl Möglichkeiten für eine Zielvariable, ihre Ausprägung zu wechseln) verstrichen ist, umso mehr sinkt die Validität eines Cues gegen Basisrate einer zufälligen Vorhersage.

Dieser Zusammenhang lässt sich experimentell implementieren, indem man Alternativen unterschiedlicher Stabilität in ein Setting einbaut und indem es einen Zeitverlauf gibt, also eine variierende Anzahl von Möglichkeiten, dass eine Alternative ihre Ausprägung ändert. Versuchspersonen sollten dann eher bei einer stabileren Alternative auf vergangene Information setzen als bei einer instabilen Alternative.

Beenden der Informationssuche. Bei der Stopppregel geht es in erster Linie um die Unterscheidung zwischen One Reason Decision Making (ökonomisches Stoppen nach dem ersten Cue) und More Reason Decision Making (Fortsetzen der Suche, auch wenn bereits eine Cue-Information vorliegt, um eine breitere Informationsgrundlage für die Entscheidung zu gewinnen). Wie es scheint, haben Individuen verhältnismässig stabile Präferenzen für das Entscheiden auf die eine oder die andere Weise (Bröder, 2000a, 2000b; Zurbriggen & Läge, 2007; Zurbriggen, Christen, Hausmann, & Läge, 2008a, 2008b).

Vor dem Hintergrund des One Reason Decision Making gewinnt die Einführung einer Zeitachse an Brisanz, denn damit liegen womöglich bereits Informationen aus vergangenen Entscheidungsdurchgängen vor und können für die aktuelle Entscheidung genutzt werden. Demnach müsste sich – zumindest bei Personen, die unter den entsprechenden Umständen dem One Reason Decision Making zugeneigt sind – eine Tendenz finden lassen, ab und zu ganz ohne neue Informationssuche zu entscheiden (was in einem Setting mit unabhängigen Einzeldurchgängen strikt vermieden wird, vgl. Hausmann, Christen & Läge, 2006). Wenn jemand aufgrund alter Informationen eine Entscheidung trifft, ohne zuvor eine neue Suche gestartet zu haben, so werden wir dies in der Auswertung der Daten als „No New Reason Stopping Rule“ (NNRStR) bezeichnen, um es der nach einem Cue gestoppten Informationssuche („One Reason Stopping Rule“, ORStR) und der auch nach einer Cue-Information noch fortgesetzten Informationssuche („More Reason Stopping Rule“, MRStR) als dritte Stopppregel zur Seite zu stellen.

Entscheidungsverhalten. Im One Reason Decision Making ist das Füllen einer Entscheidung eigentlich eine Trivialität: Man hat eine einzige diskriminierende Cue-Information auf dem

Tisch, und natürlich entscheidet man sich in diese Richtung. Im More Reason Decision Making benötigt man eine Integrationsregel, nach der die unterschiedlichen Informationen zu gewichten sind. Beide Strategien sind sowohl in statischen wie in zeitlich dynamischen Umwelten in gleicher Weise anzuwenden.

Allerdings kommt, sofern man etwas über die Stabilität der Alternativen weiss, ein weiterer Faktor hinzu. Aus der Ausprägung der Alternative in der Vergangenheit lässt sich bei hinreichender Stabilität ja auch etwas über die gegenwärtige Ausprägung sagen. Dieser Faktor wirkt allerdings nur, sofern man die tatsächliche Ausprägung der Alternative zum vorausgehenden Zeitpunkt auch kennt. Dies kann in einem Experimentalsetting durch geeignetes Vorenthalten des Feedbacks natürlich ausgeschlossen werden, sofern es im Sinne der Beantwortung der Hypothesen nicht sinnvoll ist, dass die Probanden über dieses Wissen verfügen. Doch auch wenn Personen nur eine wage Ahnung über die Ausprägung einer zeitlich stabilen Alternative haben, sollte dies ihr Vertrauen in die richtige Entscheidung bei dieser Alternative erhöhen, verglichen zu einer wenig stabilen Alternative, über die dieselbe aktuelle Information vorliegt.

Mit der Einführung einer Zeitachse ergibt sich zudem die Möglichkeit, nicht nur eine Entscheidung auf einen späteren Zeitpunkt zu verschieben, sondern auch die Informationssuche. Beides wird man vor allem dann machen, wenn die Rahmenbedingungen, überhaupt eine günstige Entscheidung treffen zu können, aus den Informationen der jüngsten Vergangenheit heraus als ungünstig beurteilt werden und wenn man sich gleichzeitig Hoffnung auf Besserung der Rahmenbedingungen in der Zukunft machen darf. Voraussetzung für eine solche Strategie ist, dass das Aufschieben keine substantiellen Ressourcen kostet.

10.2 Methode

10.2.1 Aufbau des Experimentes

Um den Unterschied zwischen einer statischen, aus Einzelentscheidungen bestehenden Umwelt und einer zeitlich dynamischen Umwelt miteinander verbundener Entscheidungssituationen experimentell zu implementieren, wurde als Information Board ein Trinkwasserszenario geschaffen: Die Versuchsperson erhält die Aufgabe, für die fiktive mexikanische Firma AquaMex Trinkwasser abzufüllen, welches jeweils zeitgleich mit zwei Tanklastwagen von zwei Quellen der Umgebung herangefahren wird. Da dieses Wasser sich nur zum Teil als Trinkwasser eignet (bei einem Drittel der Ladungen), steht neben der Abfüllstation auch eine Teststation mit vier durchführbaren Tests (den probabilistischen Hinweiscues) zur Verfügung. Diese vier Tests haben unterschiedliche Validitäten (0.90, 0.85, 0.80, 0.75). Sie können einzeln pro Wasserladung durchgeführt werden, und zwar beliebig viele in beliebiger Häufigkeit. Es ist dann zu entscheiden, ob man a) genau eine der beiden Wasserlieferungen zu Trinkwasser abfüllen möchte (die Abfüllanlage ist in ihrer Kapazität begrenzt) oder b) beide Wasserlieferungen wegschüttet (d.h. den Bauern in der Umgebung zur Bewässerung ihrer Felder zur Verfügung stellt), weil man in keine der Lieferungen genügend Vertrauen hinsichtlich der Wasserqualität hat. Entscheidet man sich für Variante a), so wird die entsprechende Lieferung in Trinkwasserflaschen abgefüllt. Eine

„Das Quellwasser wird jeweils von verschiedenen Wasserquellen geholt und mit Tanklastwagen zu AquaMex geliefert, wo es eine Wasseranalysestation und eine Abfüllstation gibt. [...] Falls du die Wasserqualität einer Tankladung für gut findest, kannst du mit dieser Tankladung zur Abfüllstation fahren. [...] Falls du der Meinung bist, dass keine der beiden Tankladungen trinkbares Quellwasser enthält, kannst du beide Lastwagen zu den benachbarten Landwirten schicken. Dies bringt dir mit Sicherheit keine Einnahmen, dafür verbrauchst du jedoch keine Abfüllzeit.“

In der zweiten Bedingung (Trinkwasser-Szenario 2, kurz T2) stammen die Tankladungen in jeder Runde aus denselben beiden Wasserquellen X und Y (siehe Abbildung 10.2). Dies impliziert die gewünschte zeitliche Abhängigkeit der Zielvariablen. Die durchgeführten Tests der letzten vier Zeitpunkte bleiben auf dem Information Board sichtbar, so dass die Versuchspersonen bei der möglichen Verwendung von Testinformationen aus der Vergangenheit nicht an einer Begrenzung des Arbeitsgedächtnisses scheitern. Die Probanden wissen über die beiden Quellen lediglich, dass sich deren Wasserqualität von Zeit zu Zeit ändern kann und dass beide im Mittel in einem Drittel der Fälle brauchbares Trinkwasser liefern. Alle anderen Umstände bleiben gleich wie in Bedingung T1. Folgender Auszug aus dem Instruktionstext für T2 zeigt den Unterschied zu T1 auf:

„In deiner Arbeit musst du dich jedoch ein wenig umstellen, denn es stehen für die Anlieferung nur noch zwei Wasserquellen zur Verfügung. Da die Wasserquellen immer dieselben zwei sind und sich die Wasserqualität nicht willkürlich ändert, bleiben die Informationen der Tests der letzten vier Zeitpunkte nutzbar. Die ankommenden Tanklastwagen bilden zwei Schlangen, die eine von Wasserquelle X kommend, die andere von Wasserquelle Y kommend. Du musst dich jeweils entscheiden, ob du eine der beiden Ladungen abfüllen willst und falls ja, welche.“

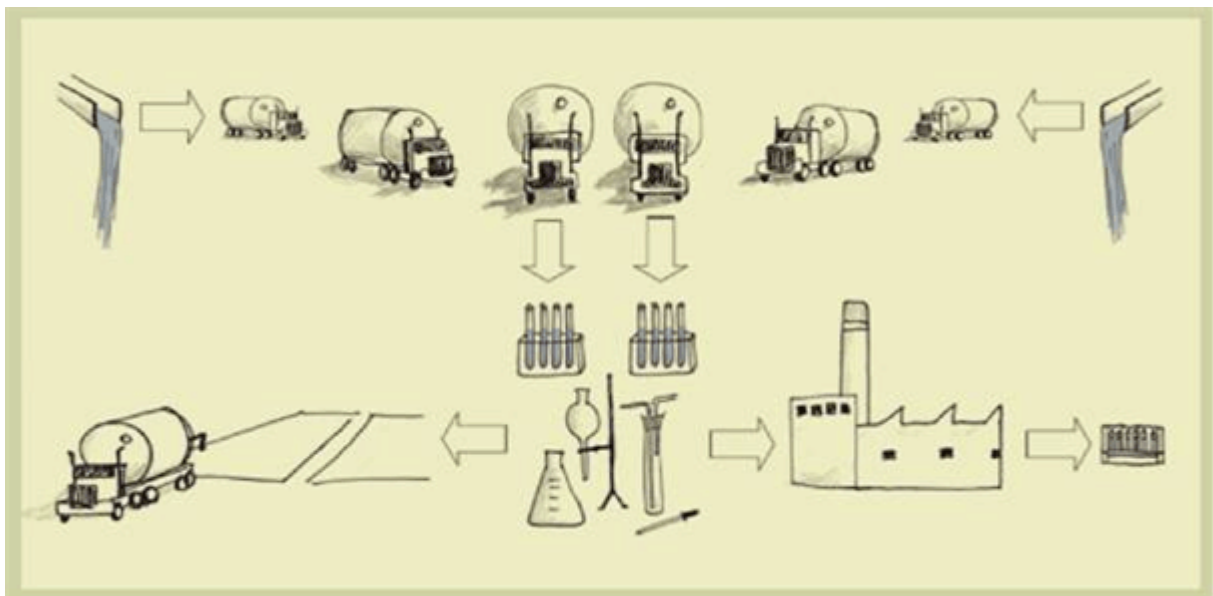


Abbildung 10.2: Illustration zur Einführung in die Bedingung des Trinkwasser-Szenarios mit zeitlich dynamischen Ereignissen (T2)

In der dritten Bedingung (Trinkwasser-Szenario 3, kurz T3) wird nun die Stabilitätsrate der Wasserqualität aus den beiden Quellen X und Y explizit mitgeteilt, ansonsten entspricht die Bedin-

gung T2. Die Stabilitätsrate der Wasserqualität ist bei den beiden Wasserquellen jedoch unterschiedlich: Quelle X hat eine Stabilität von .60 und die Quelle Y von .90.

In diesen drei Bedingungen erweitert sich jeweils das Wissen der Versuchsperson um einen Faktor. Deswegen ist es sinnvoll, die drei Bedingungen in genau dieser Reihenfolge den Versuchspersonen vorzulegen. Eine detaillierte Beschreibung aller technischen Parameter des Experiments und eine Herleitung der hinter den Prädiktor- und Zielvariablen stehenden Datenverläufe findet sich bei Christen, Zurbriggen, Hausmann & Läge (2007). Alle den Probanden mitgeteilten Validitäten, Stabilitätsraten und Basisraten sind in jeder Versuchsbedingung korrekt umgesetzt, so dass sich die Versuchspersonen auf diese Angaben verlassen können.

10.2.2 Operationalisierte Hypothesen

Hinsichtlich der Stopppregel und hinsichtlich der Entscheidungsregel erwarten wir Unterschiede im Verhalten, wenn sich die Umwelt von einer statischen zu einer zeitlich dynamischen wandelt. Allerdings enthält ein erstes Experiment mit Einführung einer entsprechenden Zeitachse natürlich einen gewissen Explorationsfaktor. Es kann z.B. nicht vorhergesagt werden kann, ob Personen allein schon durch die Möglichkeit, Informationen aus der Vergangenheit nutzen zu können, darauf reagieren werden oder doch erst mit einem Wissen über die Stabilität der Zielvariablen. Es wird also jeweils festzustellen sein, ob sich bereits ein Unterschied von T1 zu T2 ergibt oder erst von T2 zu T3.

Die erste wichtige in dieser Hinsicht zu überprüfende Variable ist die Anzahl durchgeführter Tests pro Durchgang. Mit der Möglichkeit, Informationen aus der Vergangenheit heranzuziehen, sollte sich die Zahl der aktuell durchgeführten Tests verringern (Hypothese 1).

Des weiteren eröffnet die Zeitachse die strategische Möglichkeit des Abwartens: Wenn die Tests zu beiden Quellen in jüngster Vergangenheit eine schlechte Wasserqualität angezeigt haben, dann gibt es ja die Möglichkeit, sogar ganz ohne Test beide Lieferungen den Bauern zu überlassen (Hypothese 2 = Anzahl der Durchgänge, in denen beide Lieferungen weggeschüttet werden, steigt gegenüber T1). Das ist der Extremfall der Variante. Im moderaten Fall würde vielleicht noch ein kontrollierender Test durchgeführt, bevor man die Lieferungen wegkippt. In der Summe sollte dies darin resultieren, dass durch die „gesparte“ Zeit die Anzahl der mit den 240 Stunden spielbaren Durchgänge durch die Einführung der Zeitachse zunimmt (Hypothese 3).

Das bedeutet eine Veränderung der Stopppregel: Entscheiden ohne Informationssuche (nicht zu erwarten in statischen Szenarien) sollte ab T2, spätestens jedoch bei T3 eine valable Alternative werden (Hypothese 4). Ausserdem sollte sich die Attraktivität der zusätzlichen Suche bestätigender Information über einen ersten diskriminierenden Hinweis hinaus verringern; die More Reason Stopping Rule sollte also seltener gewählt werden (Hypothese 5).

Diese Veränderungen erwarten wir allerdings nicht für die gesamte Stichprobe, sondern primär für Subgruppen, die sich anhand ihres Stoppverhaltens in T1 voneinander unterscheiden lassen: Wer in T1 mehrheitlich One Reason Decision Making betrieb, der sollte nun die Tendenz zeigen, immer einmal wieder ganz ohne neue Informationen zu entscheiden (Hypothese 6). Personen,

die in T1 zum More Reason Decision Making neigten, sollten hingegen die Cue-Informationen aus der Vergangenheit mit einbeziehen und deswegen in der jeweils aktuellen Informationssuche auch häufiger einmal mit einer One Reason Stopping Rule auskommen (Hypothese 7).

Hinter den Zielvariablen (und damit auch hinter den Prädiktorvariablen) stehen bestimmte Zeitverläufe, die den in T3 offengelegten Stabilitätsraten entsprechen. In T1 sind diese irrelevant, weil die Durchgänge logisch unverbunden sind. Mit dem Wissen um die zeitliche Verbindung in T2, spätestens aber mit dem Wissen um die Stabilitätsraten in T3, und unterstützt durch die Informationen aus der Vergangenheit, haben die Versuchspersonen die Möglichkeit, derartige Verläufe zu entdecken. Dies sollte darin resultieren, dass die Wahrscheinlichkeit dafür steigt, dass dieselbe Alternative, wenn sie beim vorausgehenden Durchgang zum Abfüllen gewählt wurde, unmittelbar wieder gewählt wird (Hypothese 8).

Mit der Einführung der Stabilitätsfaktoren in T3 wird die Quelle Y gegenüber der Quelle X attraktiver. Es steht deswegen zu erwarten, dass sich die Informationssuche in T3 im Vergleich zu T2 stärker auf Quelle Y verlagert (Hypothese 9). Dasselbe erwarten wir für die Anzahl der Entscheidungen (Hypothese 10), wobei dieser Effekt aber nicht gross sein kann, da ja beide Alternativen gleich häufig (ein Drittel der Fälle) eine gute Wasserqualität bieten. Eine Möglichkeit zu einem grösseren Effekt bietet hingegen die Betrachtung je zwei unmittelbar aufeinanderfolgender Durchgänge: Wenn die stabilere Alternative Y im vorausgehenden Durchgang als positiv bewertet wurde (also zur Abfüllung kam), dann sollte sie im aktuellen Durchgang in T3 eine höhere Wahrscheinlichkeit der Testdurchführung (Hypothese 11) und vor allem auch eine höhere Wahrscheinlichkeit der wiederholten Abfüllung (Hypothese 12) haben als noch in T2. Für die Quelle X (um deren weitaus niedrigere Stabilität die Versuchsperson ja in T3 weiss) erwarten wir das nicht.

Mit der Einführung der Stabilitätsinformation wird für die Probanden in T3 auch das Risiko kalkulierbar, wenn sie die Durchführung neuer Tests reduzieren und eher auf vergangene Information setzen. Das könnte sich im Suchverhalten und im Stoppverhalten ausdrücken, und zwar in einer Asymmetrie zwischen der stabileren Alternative (hier ist der Test vom vorausgehenden Durchgang noch mehr wert) und der weniger stabilen Alternative: Da die Versuchspersonen ja in der Regel den validesten Test durchführen dürften und dieses Resultat damit vom letzten Durchgang vorliegt, müsste es also gehäuft für die stabile Alternative Y zu Fällen kommen, in denen dieser valideste Test überhaupt nicht konsultiert wird (Hypothese 13). Personen sollten sich häufiger einmal ganz ohne aktuellen Test für die stabilere Alternative entscheiden als für die weniger stabile Alternative (Hypothese 14). Und Personen, die in T1 mehrheitliches MRDM betreiben, könnten vornehmlich bei der stabileren Alternative Testresultate aus vorherigen Durchgängen mit einbeziehen und somit mit nur einem einzigen zusätzlichen aktuellen Test genügend Sicherheit für ihre Entscheidung gewinnen (Hypothese 15).

Als Konstante über alle drei Bedingungen sollte sich die Suchregel ergeben: Hier erwarten wir keine Abweichungen von der Validitätsfolge (Hypothese 16). Auch die Leistung der Versuchspersonen (gezählt in korrekt abgefüllten Lieferungen) sollte nicht durch den Wechsel in den

Umwelten an sich steigen (denn die Zahl brauchbarer Lieferungen ist ja durch das Design gleichmässig begrenzt). Ein Zusammenhang zwischen Leistung und Bedingung könnte lediglich über eine grössere Anzahl gespielter Durchgänge moderiert sein (Hypothese 17), denn mit mehr Durchgängen erhöht sich ja die Chance des Findens zusätzlicher guter Wasserlieferungen.

Diese 17 Hypothesen sind nicht alle völlig unabhängig voneinander, sondern stellen in einigen Fällen Facetten desselben Grundphänomens dar. Fast jede von ihnen kann jedoch eintreten, ohne dass eine andere Hypothese dafür erfüllt sein müsste. Ein Hauptaugenmerk wird für die Hypothesen 1 – 8 darauf zu legen sein, ob, wenn sie eintreten, dies bereits bei T2 (nur Wissen über die Abhängigkeit) oder erst bei T3 (zusätzliches Wissen über die Stabilität) der Fall ist.

10.2.3 Versuchsdurchführung

Insgesamt absolvierten 30 Versuchspersonen (Alter $M = 29.4$, $SD = 8.0$; 14 Männer und 16 Frauen) die drei Bedingungen des Trinkwasserszenarios. Aus Gründen einer einheitlichen, aufeinander aufbauenden Instruktion wurden die drei Bedingungen jeweils in derselben Reihenfolge absolviert: zuerst die „zeitlose“ Bedingung (T1), dann die Bedingung, in welcher die Umwelt des Experiments durch die Zeitdimension erweitert wird (T2) und zuletzt die Bedingung mit Stabilitätsangaben zu den Alternativen (T3). Unter dem besten Drittel der Versuchspersonen wurde ein Gewinn von CHF 100.- verlost, um die Probanden zu einer guten Gesamtleistung zu motivieren, ohne sie aber zu extrem riskanten Grenzstrategien zu verleiten.

10.3 Resultate

Bei den im Folgenden berichteten inferenzstatistischen Auswertungen (t-Tests für abhängige Stichproben oder, wo zwischen Personengruppen getestet wurde, auch für unabhängige Stichproben) wurde $\alpha = 0.05$ mit zweiseitiger Hypothesenprüfung als Grundlage für das Akzeptieren eines Mittelwertsunterschiedes gewählt.

Leistung. Die Anzahl „guter“ Trinkwasserabfüllungen als zentrales Mass für die Leistung einer Versuchsperson im Experiment steigt von Bedingung zu Bedingung. In T1 werden im Mittel 11.8 Punkte ($SD = 2.0$, Range = 9 – 15), in T2 12.6 Punkte ($SD = 2.8$, Range = 9 – 18) und in T3 12.9 Punkte ($SD = 2.1$, Range = 9 – 16) erzielt. Die Leistungssteigerung von der ersten Bedingung (T1) zur zweiten Bedingung (T2) ist knapp signifikant ($t = 2.2$; $p < .05$), jedoch nicht signifikant ist der marginale Anstieg der Leistung von Bedingung 2 zu Bedingung 3 ($t = 0.6$; $p = .57$). Von T1 zu T2 verschlechtern sich 7 Personen in ihrer Leistung, 6 Personen stagnieren und 17 Personen verbessern ihre Leistung. Von T2 zu T3 verschlechtern sich 12 Personen in ihrer Leistung, 4 Personen halten die Leistung und 14 verbessern ihre Leistung. Die erbrachte Leistung korreliert in jeder Bedingung signifikant ($p < .01$) mit der Anzahl absolvierter Runden (in T1 $r = .69$, in T2 $r = .75$ und in T3 $r = .50$). Das heisst, dass Personen, welche eher mit Entscheidungen zuwarten, in der Tendenz bessere Entscheidungen treffen.

Verteilung der Informationssuche auf die einzelnen Cues. Die Häufigkeiten, mit welcher die Cues konsultiert werden, korrespondieren in allen drei Bedingungen mit der Reihenfolge der

Validität der vier Cues (Tabelle 10.1), und zwar gleich ob man sich die Gesamtzahl pro Bedingung oder den Durchschnitt pro Durchgang anschaut (letzterer wird mit ausgewertet, da sich die Zahl der Durchgänge über die Personen und über die Bedingungen stark voneinander unterscheiden konnte). Die Suchregel entspricht also (wie nicht anders zu erwarten) einer lexikografischen Suchstrategie, und zwar mit einem ganz klaren Sprung zwischen dem validesten und dem zweitvalidesten Cue.

Tabelle 10.1: Anzahl gesuchter Cues pro Versuchsperson in den drei Bedingungen T1 („zeitlos“), T2 (mit Zeitkontext) und T3 (mit Stabilitätsangaben zu den Alternativen) über alle Runden und pro Runde.

Alle Runden	T1	T2	T3
Cue 1 ($V=.90$)	M = 38.9 SD = 7.5	M = 42.3 SD = 15.0	M = 37.8 SD = 13.1
Cue 2 ($V=.85$)	M = 16.5 SD = 13.3	M = 17.2 SD = 16.8	M = 12.7 SD = 13.6
Cue 3 ($V=.80$)	M = 8.0 SD = 6.9	M = 6.2 SD = 7.5	M = 4.7 SD = 6.5
Cue 4 ($V=.75$)	M = 5.8 SD = 7.7	M = 4.2 SD = 7.1	M = 3.1 SD = 5.4
Summe	M = 69.2 SD = 21.6	M = 69.9 SD = 32.0	M = 58.3 SD = 28.4
Pro Runde	T1	T2	T3
Cue 1 ($V=.90$)	M = 1.5 SD = 0.3	M = 1.4 SD = 0.5	M = 1.3 SD = 0.5
Cue 2 ($V=.85$)	M = 0.7 SD = 0.6	M = 0.6 SD = 0.6	M = 0.5 SD = 0.5
Cue 3 ($V=.80$)	M = 0.3 SD = 0.3	M = 0.2 SD = 0.3	M = 0.2 SD = 0.3
Cue 4 ($V=.75$)	M = 0.3 SD = 0.4	M = 0.2 SD = 0.3	M = 0.1 SD = 0.2
Summe	M = 2.8 SD = 1.2	M = 2.3 SD = 1.2	M = 2.0 SD = 1.1

Bemerkenswert sind sicherlich die ab Cue 2 stark ansteigenden Standardabweichungen. Dies weist darauf hin, dass sich zwei Untergruppen von Versuchspersonen bilden, die einen, die nach dem validesten Cue die Suche abbrechen, und die anderen, die danach noch regelmässig weiter-suchen. Auch die stark ansteigende Standardabweichung von Cue 1 von Bedingung 1 zu Bedingung 2 ist bemerkenswert. Offenbar weitet sich mit dem Wechsel zu einer Umwelt mit Zeitachse die Verhaltensvarianz der Versuchspersonen hinsichtlich des Durchführens des ersten Tests enorm aus. Wir werden zu einem späteren Zeitpunkt der Auswertung feststellen, dass dies damit zu tun hat, dass eine Reihe von Personen in einer substantiellen Zahl von Fällen ganz ohne Informationssuche entscheidet (also weniger Cue 1 konsultiert) und dass sich eine andere Gruppe noch stärker als zuvor auf Cue 1 konzentriert.

Ein extrem seltenes Ereignis ist das Ignorieren des validesten Cues in einem Durchgang: Wenn Information gesucht wird, dann ist er auch praktisch immer mit dabei. Mit Blick auf die aufgestellte Hypothese 13 ist für diese Fälle interessant, wie häufig sich die Probanden in T3 mit Tests, aber ohne Cue 1 für eine der beiden Alternativen entschieden haben. Insgesamt handelt es

sich dabei um 40 Fälle, verteilt auf acht Versuchspersonen: 25mal fiel dabei die Entscheidung auf die stabilere Alternative Y ($M = 3.1$, $SD = 3.8$), 15mal auf die weniger stabile Alternative X ($M = 1.9$, $SD = 1.9$). Im t-Test wird dieser Unterschied nicht signifikant ($t = 1.1$, $p = 0.29$), wobei ein signifikanter Unterschied im Blick auf die nur acht Probanden, die sich überhaupt in Umgehung von Cue 1 für das Abfüllen von Wasser entscheiden, statistisch ohnehin problematisch wäre.

Anzahl durchgeführter Tests. Insgesamt werden in der „zeitlosen“ Bedingung T1 in etwa gleich viele Informationen wie in der zweiten Bedingung mit Zeitdimension T2 gesucht ($MT1 = 69.2$; $MT2 = 69.9$; $t = 0.2$; $p = .83$). Kommen bei der dynamischen Umwelt jedoch noch zusätzlich Stabilitätsinformationen hinzu (T3), werden signifikant weniger Informationen gesucht ($MT3 = 58.3$; $t = 5.0$; $p < .01$). Diese Abnahme ermöglicht den Versuchspersonen, im Mittel einmal mehr Wasser abzufüllen (Zeitbedarf 10 Stunden) als in T1 und T2. Pro Runde betrachtet – und das ist im Sinne der Hypothesen der wichtigere Wert – sinkt die Zahl der durchgeführten Tests von $MT1 = 2.8$ über $MT2 = 2.3$ auf $MT3 = 2.0$. Die Unterschiede zwischen je zwei aufeinander folgenden Bedingungen sind signifikant ($t_{T1-T2} = 4.1$, $p < .01$; $t_{T2-T3} = 4.0$, $p < .01$).

In allen drei Bedingungen wird häufiger bei der Alternative X als bei der Alternative Y nach Information gesucht. In T1 werden 1234 Tests zur Alternative X durchgeführt und nur 841 zur Alternative Y ($t = 5.9$; $p < .01$). In T2 wird 1113mal bei der Alternative X und 981mal bei der Alternative Y ein Test konsultiert ($t = 3.7$; $p < .01$). In T3 allerdings verschwindet der Unterschied fast, mit 890 Tests zu Alternative X und 860 Tests zu Alternative Y ($t = 0.4$; $p = .69$).

Tabelle 10.2: Häufigkeit der Suche bei derjenigen Alternative, welche in der unmittelbar vorhergehenden Runde (Zeitpunkt $t-1$) gewählt wurde. Ausgewertet über alle Runden.

	Alter-native	Total Wahl	Total Suche, wenn Alternative zu (t-1) gewählt wurde	Anteil
T1	X	342	268	0.78
T1	Y	162	51	0.31
T2	X	256	201	0.79
T2	Y	237	96	0.41
T3	X	258	184	0.71
T3	Y	276	171	0.62

Tabelle 10.2 zeigt die Wahrscheinlichkeit, mit der zu einer Alternative ein Test durchgeführt wird, wenn in der unmittelbar vorhergehenden Runde diese Tankladung für das Abfüllen des Wassers gewählt wurde. Diese Wahrscheinlichkeit ist für Alternative X (0.78, 0.79, 0.71) jeweils deutlich höher als für Alternative Y (0.31, 0.41, 0.62). Allerdings verschwindet diese Asymmetrie in T3 zum grössten Teil, indem der Wert für Quelle Y von 0.41 (in T2) auf 0.62 (T3) drastisch ansteigt.

Stoppen der Informationssuche. Wie wir bei der Tabelle 10.1 bereits gesehen haben, werden mit zunehmenden Stabilitätsangaben weniger Cues pro Runde gesucht. Zusätzlich zeigt sich, dass auch mehr Runden absolviert werden, wenn man etwas über die Stabilität der Umwelt weiss: In

T1 werden im Durchschnitt 26.6 Runden (SD = 5.0) absolviert, in T2 sind es 31.2 Runden (SD = 5.3) und in T3 31.5 Runden (SD = 9.5). Die Anzahl Runden unterscheidet sich zwischen den ersten beiden Bedingungen signifikant, zwischen den zeitlich dynamischen Bedingungen hingegen nicht mehr ($t_{T1-T2} = 8.80$; $p < .01$; $t_{T2-T3} = 0.19$, $p = .86$).

Bei der Einführung der Hypothesen haben wir in Anlehnung an die bereits bestehende Forschung eine Klassifizierung des Stoppverhaltens in drei Kategorien vorgeschlagen: No-New-Reason Stopping Rule (Entscheiden ohne Hinzuziehen von neuen Informationen), One-Reason Stopping Rule und More-Reason Stopping Rule. Da im vorliegenden Experiment die Informationen je Alternative einzeln gesucht wurden, sind die möglichen resultierenden Fälle des Abbruchs einer Informationssuche vor der Auswertung der Daten diesen drei Stoppregeln zuzuordnen. Es gilt:

- Keine Information in einem Durchgang gesucht = NNRStR
- Stoppen nach dem ersten angeschauten Cue, gleich ob dieser positiv oder negativ ist = ORStR
- Fortsetzen der Informationssuche mit der anderen, nachdem der erste angeschaute Cue zur einen Alternative negativ war, und Stoppen nach einem angeschauten Cue der anderen Alternative (gleich ob dieser positiv oder negativ ist) = ORStR
- Suchen von mehr als einem Cue zur selben Alternative = MRStR
- Suchen nach einem Cue zur anderen Alternative, nachdem zur einen Alternative bereits ein positiver Cue gefunden wurde = MRStR

Auf der Mittelwertsebene zeigt Tabelle 10.3 (unterste Spalte *Sum*) eine Tendenz zu häufigerem Anwenden von ORStR in allen drei Bedingungen. In der ersten „zeitlosen“ Bedingung T1 stoppen die Probanden insgesamt etwas häufiger gemäss einer ORStR ($M = 13.8$, $SD = 12.4$), als gemäss einer MRStR ($M = 12.4$, $SD = 9.3$), jedoch nicht signifikant ($t = 0.4$; $p = .73$). Zudem entscheiden sich die Versuchspersonen in T1 wie angenommen kaum je ohne eine Information ($M = 0.3$, $SD = .61$; sieben der zehn Fälle finden in der allerersten oder allerletzten Runde statt). In T2 wird ebenfalls häufiger anhand einer einzigen Information entschieden ($M = 14.5$, $SD = 11.5$) als anhand mehrerer Informationen ($M = 12.6$, $SD = 10.8$), aber auch hier nicht signifikant ($t = 0.5$; $p = .64$). Dafür entscheiden sich die Probanden in T2 in vielen Runden ohne aktuelle Informationssuche ($M = 4.0$, $SD = 7.9$). In T3 wird wiederum häufiger anhand einer einzigen Information entschieden ($M = 15.8$, $SD = 11.2$) als anhand mehrerer ($M = 10.4$, $SD = 9.9$), aber auch das nicht signifikant ($t = 1.5$; $p = .15$). Häufig wird auch hier ohne jegliche Information entschieden ($M = 5.2$, $SD = 9.8$).

Die Anzahl der Fälle, in welchen ohne eine neue Information entschieden wird, unterscheiden sich in T1 und T2 signifikant ($t = 2.5$; $p < .05$). Jedoch wird in beiden Bedingungen gleich häufig gemäss ORStR ($t = 0.4$; $p = .66$) gestoppt und auch gleich häufig gemäss MRStR ($t = 0.3$; $p = .74$). Die Häufigkeiten der Stoppregeln in den beiden dynamischen Bedingungen T2 und T3 unterscheiden sich weder in der Anzahl Fälle, in welchen keine Information gesucht wird ($t =$

1.1; $p = .28$), noch in Bezug auf die ORStR ($T = 1.0$; $p = .31$). Jedoch sinkt die Häufigkeit des MRStR von T2 nach T3 hin signifikant ($T = 2.3$; $p < .05$).

Tabelle 10.3: Anzahl Runden, in denen eine Versuchsperson a) ohne neue Information eine Entscheidung trifft (NNRStR=N) b) sich mit einer Information entschied (ORStR=O) oder c) bei bereits vorliegender Information noch weiter sucht (MRStR=M). In den Spalten sind die 30 Versuchspersonen angegeben, sortiert nach absteigender Differenz zwischen ORStR und MRStR in Bedingung T1. Die Majorität des Verhaltens legt jeweils die Strategie dieser Person (fett gedruckt) fest.

Vp	T1			T2			T3		
	N	O	M	N	O	M	N	O	M
14	2	45	0	24	26	1	42	23	0
29	0	29	0	22	13	0	20	13	0
5	0	28	0	2	29	0	0	29	0
11	0	28	0	22	7	0	17	11	0
23	1	28	0	0	29	1	4	23	0
22	2	27	0	1	29	0	0	29	0
10	0	26	2	0	34	2	0	25	5
25	0	26	2	5	28	2	24	29	0
3	0	26	3	0	31	1	15	35	1
8	0	24	3	0	28	3	0	44	0
4	0	21	7	24	4	7	12	16	1
21	0	18	9	0	28	3	0	9	20
27	0	15	11	6	18	8	0	15	10
24	0	15	12	11	18	7	12	16	0
17	0	12	17	0	19	14	0	20	10
12	0	7	15	2	9	11	1	14	8
6	1	7	16	0	2	25	0	14	11
13	0	7	18	0	23	11	0	29	0
16	0	9	20	0	12	21	0	19	14
18	1	5	18	0	8	17	0	19	11
28	1	0	16	1	0	22	3	0	22
26	0	4	23	0	11	21	0	6	22
2	0	1	21	0	12	19	1	9	22
15	0	1	21	0	1	19	5	1	16
1	0	3	24	0	0	30	1	5	23
7	1	0	21	0	0	29	0	1	25
20	1	0	21	0	9	21	1	5	22
9	0	1	23	0	6	28	0	9	20
19	0	0	24	0	1	28	0	0	27
30	0	1	26	0	0	29	0	6	22
Sum	10	414	373	120	435	380	158	474	312

Das Stoppverhalten auf individueller Ebene steht in Tabelle 10.3 im Mittelpunkt. Dabei wurde die jeweils dominante Stoppstrategie einer Person gemäss Majoritätsregel festgelegt (fett eingefärbt), ist also nicht per Binomialtest abgesichert und dient deswegen der deskriptiven Darstellung und nicht der Auszählung klar inferenzstatistisch identifizierbarer Strategien. In T1 werden danach 14 Personen als ORStR-Anwender klassifiziert und 16 Personen stoppen mehrheitlich gemäss einer MRStR. 19 der 30 Probanden behalten ihre Majoritäts-Stoppregel über das gesamte Experiment hin bei (10 aus der MRStR-Gruppe und 9 aus der ORStR-Gruppe). Vier der 10 aus der MRStR-Gruppe stoppen jedoch in T2 und T3 markant häufiger auch einmal nach einem einzigen Cue, während eine umgekehrte Tendenz bei den 10 konstanten der ORStR-Gruppe angehörenden Personen nicht zu erkennen ist. Im Gegenteil, sie entscheiden häufiger einmal ganz

ohne Informationen (zwei Personen in T2 und fünf in T3), was sich aus der MRStR-Gruppe lediglich Vp15 in T3 ein paar Mal erlaubt.

Bei den Majoritätswechslern setzt sich dieser Trend fort, und zwar intraindividuell in entsprechend verstärkter Form (Strategiewechsel eben): Sechs ursprünglich die MRStR bevorzugende Personen wechseln entweder gleich in T2 oder spätestens in T3 zur ORStR, während vier Personen, die in T3 die ORStR bevorzugten, mit der Einführung der Zeitachse zur NNRStR als Majoritätsfall wechseln (Vp14 allerdings in T2 erst halbherzig). Die Tendenz ist also klar: Personen, die in einer Umwelt ohne Zeitachse die MRStR bevorzugten, zeigen bei der Möglichkeit, auch Informationen aus der Vergangenheit heranzuziehen, mehrheitlich eine Tendenz, mit einer neu gesuchten Information zufrieden zu sein. Zur NNRStR hingegen wechselt niemand aus dieser Gruppe. Bei vielen Personen, welche in einer Umwelt ohne Zeitachse zum One Reason Decision Making tendierten, zeigt sich eine klare Tendenz, nun häufiger einmal ganz auf die Informationssuche zu verzichten und lediglich auf die Information aus der Vergangenheit zu vertrauen. Keine Regel ohne Ausnahme allerdings: Vp21, die mit One Reason Decision Making startet, verwendet in T3 schliesslich doch die MRStR.

In den 158 Durchgängen, in denen in T3 ohne eine Information entschieden wurde, wurde 81mal eine der beiden Wasserlieferungen abgefüllt. Dabei wurde 35mal die Alternative X und 46mal die Alternative Y gewählt. Vergleicht man diese Werte bei den zwölf Probanden, die für mindestens eine der beiden Alternative einen Score >0 haben, so ergibt sich kein signifikanter Unterschied bei den beiden Alternativen ($t = 0.7$, $p = .49$).

Für die 16 Probanden, die in T1 mehrheitlich MRDM gezeigt hatten, sind die Anzahl Fälle bei den beiden Alternativen in T3 zu vergleichen, in denen anhand einer einzigen neuen Information für eine Alternative entschieden wurde. Dies ist bei der Alternative X 135mal der Fall und bei der Alternative Y 165mal. Dieser Unterschied ist nicht signifikant ($t = 0.2$, $p = .86$).

Tabelle 10.4: Häufigkeit der Wahl der drei möglichen Entscheidungsalternativen: Summen über alle Probanden, individuelle Mittelwerte und Standardabweichungen für alle drei Experimentalbedingungen.

	T1	T2	T3
Alternative X	342	256	258
	M = 11.4	M = 8.8	M = 8.6
	SD = 2.7	SD = 2.4	SD = 4.6
Alternative Y	162	237	276
	M = 5.4	M = 7.9	M = 9.2
	SD = 2.3	SD = 2.9	SD = 4.7
Felder bewässern	293	433	410
	M = 9.7	M = 14.4	M = 13.7
	SD = 4.1	SD = 5.2	SD = 9.0

Entscheidungsverhalten. Wie Tabelle 10.4 zeigt, wird in T1 die Alternative X deutlich häufiger gewählt als die Alternative Y ($t = 7.2$; $p < .01$). Diese extreme Asymmetrie verschwindet in T2 weitgehend ($t = 1.2$; $p = .24$) und kippt in T3 sogar (jedoch nicht signifikant; $t = 0.4$; $p = .72$).

Der entscheidende Wechsel im Entscheidungsverhalten findet zwischen T1 und T2 statt: Alle drei Wahlhäufigkeiten, die der Alternative X die der Alternative Y und die des Felderbewässerns, unterscheiden sich zwischen T1 und T2 signifikant voneinander: Alternative X wird weniger gewählt ($t_{T1-T2} = 5.3$, $p < .01$), Alternative Y dafür häufiger ($t_{T1-T2} = 5.0$, $p < .01$), und ebenfalls häufiger entscheiden sich die Personen dafür, beide Wasserladungen wegzuschütten ($t_{T1-T2} = 7.0$, $p < .01$). Zwischen T2 und T3 hingegen gibt es keinerlei signifikanten Veränderungen ($t_{T2-T3} = 0.3$; $p = .79$ bei Alternative X, $t_{T2-T3} = 1.4$; $p = .18$ bei Alternative Y und $t_{T2-T3} = 0.6$; $p = .57$ beim Wegschütten beider Ladungen).

Tabelle 10.5: Häufigkeit der Wahl derselben Alternative in der unmittelbar nächsten Runde, aufsummiert über alle 30 Probanden.

	Alternative	Total Wahl	Total Wiederwahl	Anteil
T1	X	342	176	0.51
T1	Y	162	61	0.38
T2	X	256	166	0.65
T2	Y	237	137	0.58
T3	X	258	161	0.62
T3	Y	276	198	0.72

Tabelle 10.5 zeigt die Häufigkeit der unmittelbaren Wiederholung der Wahl einer Alternative in zwei aufeinander folgenden Durchgängen. Diese ist zunächst in T1 für Alternative X höher als für Alternative Y (was schlicht an der grösseren Basishäufigkeit der totalen Wahl liegen dürfte). Dieser Effekt reduziert sich in T2, da der Unterschied in der Basishäufigkeit verschwindet.

Das eigentlich zu bemerkende Resultat ist jedoch der markante Anstieg beider Wahrscheinlichkeiten (auch die der Alternative X, obwohl die Basishäufigkeit sinkt). Die Versuchspersonen scheinen also auf die Änderung in der Umwelt zu reagieren (und dies, obwohl die hinter dem Experiment liegenden Daten dieselben waren), die zeitlichen Verläufe der guten und schlechten Wasserqualität zu entdecken und entsprechend häufiger auf die Wahl einer Alternative die Wahl derselben Alternative im nächsten Durchgang folgen zu lassen. Diese Tendenz nimmt in T3 weiter zu allerdings nur für die (wesentlich stabilere) Alternative Y.

10.4 Diskussion

Die doch recht umfassende Darstellung der Ergebnisse verlangt zunächst nach einer Zusammenfassung im Sinne der 17 oben aufgestellten Hypothesen. Wir beginnen mit den „Routineüberprüfungen“ Validitätsfolge und Leistung: Die Häufigkeiten der durchgeführten Tests entsprechen in allen drei Bedingungen (wie nicht anders zu erwarten) der Suchregel in absteigender Validitätsfolge (Hypothese 16). Die Leistungen der Versuchspersonen unterscheiden sich zwischen den drei Bedingungen nicht stark voneinander, nehmen allerdings von 11.8 erfolgreichen Wasserabfüllungen in T1 über 12.6 in T2 bis zu 12.9 in T3 mit zunehmender Anzahl gespielter Durchgänge leicht zu (weil damit die Chancen steigen, zusätzliche gute Wasserlieferungen zu finden). Innerhalb jeder Bedingung erweisen sich die Personen, die mehr Durchgänge mit ihrem Zeitbud-

get spielen konnten, als die besseren (Korrelationen jeweils $p < .01$ signifikant), so dass wir den gerade bei $p < .05$ signifikanten Unterschied zwischen T1 und T2 ebenfalls diesem Effekt zunehmender Spieldurchgänge zuschreiben können (Hypothese 17).

Für die Hypothesen 1 – 8 ist nicht nur festzustellen, ob sie eintraten oder nicht, sondern es ist gleichermassen interessant zu sehen, ob dafür die reine Einführung der zeitlich dynamischen Umwelt ausreicht oder ob die Versuchspersonen für die erwartete Verhaltensänderung auf das Wissen über den Grad an Stabilität der Zielvariablen angewiesen sind.

Die erste in dieser Hinsicht zu überprüfende Variable ist die Anzahl durchgeführter Tests pro Durchgang (Hypothese 1). Wie erwartet, verringert sich diese Anzahl signifikant, und zwar bereits von T1 (im Mittel 2.8 Tests) zu T2 (2.3 Tests). Von T2 zu T3 (2.0 Tests) ist dann noch einmal eine signifikante Verringerung festzustellen. Das zusätzliche Wissen um die Stabilität erlaubt den Versuchspersonen also, ihre Tests noch gezielter einzusetzen.

Die Anzahl der Durchgänge, in denen beide Lieferungen weggeschüttet werden (Hypothese 2), steigt gegenüber T1 ($M_{T1} = 9.7$) bereits in T2 massiv an ($M_{T2} = 14.4$), um dann in T3 in etwa auf diesem Niveau zu verbleiben ($M_{T3} = 13.7$). Auch die Anzahl gespielter Runden (Hypothese 3) steigt entsprechend von T1 ($M_{T1} = 26.6$) zu T2 hin massiv an ($M_{T2} = 31.2$) und bleibt dann in T3 auf diesem Level ($M_{T3} = 31.5$). Die Versuchspersonen ergreifen die strategische Möglichkeit des Abwartens, die durch die Zeitachse ermöglicht wird, also bereits unmittelbar. Das Wissen um die Stabilität der Zielvariablen hat hierauf keinen Einfluss.

Mit dem Einsatz dieser Strategie passt sich das Stoppverhalten entsprechend an: War das Entscheiden ganz ohne Information in der statischen Umwelt T1 keine Option ($M_{T1} = 0.3$), so wird dies zu einer, sobald in T2 auf bereits vorliegende Informationen aus der Vergangenheit zurückgegriffen werden kann ($M_{T2} = 4.0$). In T3 ($M_{T3} = 5.2$) steigt dieser Wert noch einmal leicht, aber nicht signifikant an. Auch Hypothese 4 kann also bestätigt werden, und zwar wie Hypothese 2 und 3 bereits im Wechsel von T1 zu T2. Erst später (von T2 zu T3) bestätigt sich allerdings Hypothese 5, die eine Abnahme des More Reason Decision Making vorhersagte ($M_{T2} = 12.6$, $M_{T3} = 10.4$; $p < .05$). Es sind die von T1 zu T2 zusätzlich gewonnenen Durchgänge, die – zumindest auf der Ebene der gesamten Stichprobe – in die Erhöhung der Entscheidungen ohne aktuelle Informationssuche gesteckt wird (wie gezeigt). Die MRStR hingegen bleibt in ihrer Häufigkeit von der ersten Bedingung zur zweiten konstant ($M_{T1} = 12.4$).

Differenzierter wird dieses Bild mit einer Betrachtung der individuellen Personen bzw. der Subgruppen, die wir aufgrund des individuellen Stoppverhaltens in T1 gebildet haben: In der Tat sind es die Personen, die in T1 mehrheitlich One Reason Decision Making betrieben, welche die Tendenz zeigen, immer einmal wieder ganz ohne neue Informationen zu entscheiden (Hypothese 6, im Resultateteil in Verbindung mit Tabelle 10.3 im Detail erläutert). Wer hingegen in T1 zum More Reason Decision Making neigt, der zeigt die Tendenz, die Cue-Informationen aus der Vergangenheit mit zu berücksichtigen und kommt deswegen in der jeweils aktuellen Informationssuche häufiger einmal mit einer One Reason Stopping Rule aus (Hypothese 7, ebenfalls detailliert in Verbindung mit Tabelle 10.3 ausgeführt). Allerdings haben wir bei der Besprechung die-

ser Resultate gesehen, dass nur jeweils ein Teil (12 von 16 Personen von der MRStR zur ORStR sowie 8 von 14 Personen von der ORStR zur NNRStR) dieser Tendenz folgt, einige allerdings so intensiv, dass sich dadurch ihre Majoritätsstrategie verschiebt. Insgesamt reagieren somit 20 der 30 Versuchspersonen (67%) mit einer angepassten Stopppregel auf die Veränderung der Umwelt, ein Drittel reagiert nicht darauf und behält seine Stopppregel bei (bzw. ändert sie im Fall von Vp21 in die nicht erwartete Richtung). Nicht eindeutig auszumachen ist eine Reaktion bereits in T2 oder erst zu T3. Hier vermittelt die Gesamtstichprobe ein uneinheitliches Bild (11 der 20 Personen verändern ihr Stoppverhalten in T2, 9 in T3).

In Hypothese 8 hatten wir uns darauf bezogen, dass im gesamten Experiment hinter den Zielvariablen (und damit auch hinter den Prädiktorvariablen) bestimmte Zeitverläufe stehen, die allerdings in T1 für die Versuchspersonen irrelevant sind (da sie ja unabhängige Durchgänge spielen) und in T2 noch völlig intransparent. Erst mit den Stabilitätsangaben in T3 erhalten die Probanden eine Ahnung darüber, wie diese Zeitverläufe aussehen können. Allerdings wissen sie dadurch natürlich noch nicht, wie sich diese Verläufe über das Experiment verteilen. Doch das können sie mit der Informationssuche ja herausfinden. Dies sollte, so unsere Erwartung, darin resultieren, dass die Wahrscheinlichkeit dafür steigt, dass dieselbe Alternative, wenn sie beim vorausgehenden Durchgang zum Abfüllen gewählt wurde, unmittelbar wieder gewählt wird. Wegen der starken Asymmetrie der gewählten Alternativen (die wir später noch zu diskutieren haben) sind die Werte in Tabelle 10.5 nach Alternativen aufgeschlüsselt. Man sieht darin aber sehr deutlich, dass der generelle Anstieg (gleich für welche Alternative) von T1 zu T2 erfolgt: Für Alternative X von 0.51 auf 0.65 und für Alternative Y von 0.38 auf 0.58. Bereits die Einführung einer Zeitachse reicht also aus, damit die Personen mit Hilfe der Informationssuche die zeitlichen Verlaufscharakteristiken der Zielvariablen erkennen. Ein Wissen über den Stabilitätsgrad der Zielvariablen ist dazu nicht notwendig.

Dieses Wissen hat allerdings trotzdem noch Auswirkungen, wie die Überprüfung der Hypothesen 9 – 12 zeigt. Die Informationssuche verlagert sich in T3 im Vergleich zu T2 stärker auf Quelle Y (Hypothese 9): In T2 bestand die noch zu diskutierende Asymmetrie (1113 Tests für Alternative X, 981 Tests für Alternative Y, $p < .05$), in T3 ist dieser signifikante Unterschied dann praktisch vollständig aufgehoben (890 Tests für Alternative X, 860 Tests für Alternative Y, $p = .81$). Dasselbe gilt – zumindest als Tendenz – auch für die Anzahl der Entscheidungen (Hypothese 10): Wie Tabelle 10.4 zeigt, entscheiden sich in T2 die Personen häufiger für Alternative X (256mal) als für Alternative Y (237mal). Während der Wert für Alternative X in T3 praktisch unverändert bleibt (258mal), wird Alternative Y attraktiver (276mal) und überflügelt Alternative X. Allerdings verfehlt der Anstieg der Attraktivität das Kriterium der Signifikanz ($p = .18$). Bereits beim Aufstellen von Hypothese 10 war freilich kein grosser Effekt erwartet worden, da ja beide Alternativen gleich häufig (ein Drittel der Fälle) eine gute Wasserqualität bieten.

Weitaus deutlicher wird dieser Effekt, wenn man das Verhalten bei direkt zuvor abgefüllter Alternative Y bzw. Alternative X betrachtet: Wenn die stabilere Alternative Y im vorausgehenden Durchgang als positiv bewertet wurde (also zur Abfüllung kam), dann erhöht sich im aktuellen Durchgang in T3 die Wahrscheinlichkeit der Testdurchführung von 58% (in T2) auf 72% (vgl.

Tabelle 10.5). Das bestätigt Hypothese 11. Gleichzeitig steigt auch die Wahrscheinlichkeit der wiederholten Abfüllung an, und zwar von 41% auf 62% (siehe Tabelle 10.2). Auch Hypothese 12 bestätigt sich also. Ein Anstieg wie bei der stabilen Alternative Y geschieht bei der weniger stabilen Alternative X nicht. Im Gegenteil, die Wahrscheinlichkeit der wiederholten Abfüllung sinkt sogar leicht (von 65% auf 62%), genauso wie die Wahrscheinlichkeit der Informationssuche direkt nach einer Abfüllung (von 79% auf 71%).

Ihre Grenze findet die Verwendung der Stabilitätsinformation allerdings beim Vertrauen auf zeitlich vorausliegende Tests. Die Resultate zu den Hypothesen 13 bis 15 ergeben zwar für T3 alle eine Tendenz zugunsten der stabileren Alternative Y, jedoch wird keiner der t-Tests signifikant: Personen umgehen bei Entscheidungen für Alternative Y nur leicht häufiger den validesten Cue (25mal gegenüber 15mal für Alternative X), sie tendieren nur etwas häufiger dazu, ganz ohne neuen Test für Alternative Y zu entscheiden (46mal gegenüber 35mal für Alternative X), und die 16 More-Reason-Decision-Maker aus T1 vertrauen in T3 nur etwas häufiger bei einer Entscheidung zugunsten von Alternative Y auf einen einzigen neuen Test (165mal gegenüber 135mal für Alternative X). Die Tendenz geht zwar jedes Mal in die erwartete Richtung, signifikant wird dieser Unterschied aber zu keiner der drei Hypothesen. Wir stellen somit fest, dass die Stabilitätsinformation über die Alternativen das Vertrauen in Information aus der Vergangenheit in diesem Experiment nicht systematisch erhöht.

In der Summe bestätigen die Resultate jedoch fast alle aufgestellten Hypothesen. Die Versuchspersonen reagieren auf die veränderte Umwelt, und zwar in der Regel bereits direkt mit der Einführung einer Zeitachse. Lediglich bei der Anpassung der Stopppregel teilt sich die Stichprobe in solche Personen auf, die unmittelbar in T2 reagieren, in solche, die erst in T3 reagieren, und in solche, welche überhaupt nicht auf die Veränderung der Umwelt reagieren. Dass diese drei Subgruppen praktisch gleich gross sind, sollte aufgrund der begrenzt grossen Stichprobe zunächst einmal als eine Zufälligkeit in den Resultaten betrachtet werden und kann nicht als ein quantitatives Ergebnis dieser Studie herhalten. Was vor allem das letzte Drittel deutlich macht ist, dass die Reaktion auf die Veränderung der Umwelt keine Selbstverständlichkeit ist. Vermutlich decken die signifikanten Mittelwertsunterschiede zu, dass sich auch bei denjenigen Massen, die für die Hypothesentestung nur auf Mittelwertebene geprüft wurden, längst nicht alle Versuchspersonen in die jeweils gefundene Richtung verhalten.

Besonders zu diskutieren ist die enorme Asymmetrie, die sich vor allem in Bedingung T1 im Suchverhalten zwischen Alternative X und Alternative Y zeigt: Für Alternative X werden in der Regel mehr Tests durchgeführt als für Alternative Y, und beim positiven Ausgang wird dann auch Alternative X massiv häufiger gewählt. Diese Asymmetrie dürfte auf die Gewohnheit zurückzuführen sein, auf einem Bildschirm (oder auf einem Blatt Papier) oben links zu beginnen (sofern es keinen Grund gibt, an anderer Stelle einzusteigen). Und links steht eben immer Alternative X. Insofern ist diese Asymmetrie für die Beantwortung der Fragestellungen irrelevant. Ja sie stört nicht einmal wirklich, denn dadurch wird – unvorhergesehenerweise – in Bedingung 1 eine Befundlage geschaffen, die sich in Bedingung 2 und 3 replizieren sollte, wenn die Versuchspersonen keinen markanten Unterschied in den Bedingungen sehen und damit keinen

Grund, an anderer Stelle als oben links mit einem jeden Durchgang zu beginnen. Die Asymmetrie reduziert sich aber in Bedingung 2 markant und verschwindet ganz in Bedingung 3. Das ist also ein zusätzlicher Hinweis darauf, dass die Versuchspersonen auf die geänderte Umwelt eingehen.

10.5 Schlusdiskussion

Die Resultate zeigen, dass Personen in dem experimentellen Setting ihr Verhalten eindeutig an die neu eingeführte Zeitachse anpassen. Zusammenfassend kann gesagt werden, dass sie die Informationssuche im zeitlichen Kontext reduzieren, sich ab und zu auch ohne eine aktuelle Information entscheiden und häufiger auch das „Nichtstun“ wählen. Die Personen beziehen somit vergangene Informationen mit in die Planung der anstehenden Informationssuche ein und optimieren so ihr Suchverhalten in Bezug auf das Kosten-Nutzen-Verhältnis. Die mit der Zeitdimension einher gehende Stabilität wird zum Zeitsparen genutzt und erbringt damit einen prospektiven Gewinn.

Wird eine Zeitachse eingeführt, treten zudem vermehrt Wiederholungen bei der Suche und bei der Wahl einer Alternative auf. Dieses Verhalten ist sinnvoll, da in der zweiten Bedingung T2 immer dieselben beiden Quellen zur Auswahl stehen. Dies verleitet die Versuchspersonen möglicherweise zur Annahme, dass die Zielvariablen über eine gewisse Stabilität verfügen. Sie liegen damit auch richtig, müssen jedoch das Pattern durch aktive Informationssuche entdecken.

Das Experimentaldesign hat somit zu einer Reihe von ersten Erkenntnissen zur Implementierung einer Zeitachse und zur Implementierung der Ressource Zeit im Informationssuch- und Entscheidungsverhalten geführt. Gleichzeitig war das Design jedoch in der Kosten-Nutzen-Struktur etwas asymmetrisch aufgebaut, indem das Wegschütten von Wasser keinerlei Ressourcen benötigte. Versuchspersonen hätten also ab T2 auf die Strategie kommen können, bei negativen Testinformationen zu beiden Wasserlieferungen einfach ohne weitere Tests eine ganze Reihe von Wasserlieferungen einfach wegzuschütten und erst dann wieder zu testen, ob sich die Situation gebessert hätte. Diese Strategie (welche lediglich von Vp14 und von Vp25 gehäuft angewendet wurde), ist ein Zeichen der vollständigen Einsicht in die statistische Struktur der Umwelt, jedoch „verwässert“ sie gleichzeitig die Auswertung im Hinblick auf die meisten der 17 operationalisierten Hypothesen. Insofern waren wir froh, dass lediglich zwei Probanden diese Strategie verwendet haben. Für zukünftige Experimente mit diesem Trinkwasser-Design würden wir jedoch überlegen, ob es nicht günstiger wäre, auch für das Wegschütten einer Wasserladung jeweils einen Aufwand (z.B. 1 Stunde pro Lieferung) anzusetzen.

Ob die Festlegung auf eine fixe Reihenfolge der drei Experimentalbedingungen eine Limitation für die Resultate darstellt oder nicht, kann kontrovers diskutiert werden. Auf der einen Seite ist es guter Brauch, im Falle mehrerer Bedingungen die Reihenfolge ausgewogen oder mindestens doch randomisiert zu halten. Das setzt jedoch eine gewisse Austauschbarkeit der Bedingungen voraus. Hinsichtlich des hier eingeführten Szenarios erscheint uns diese Austauschbarkeit jedoch nicht gegeben zu sein: Wenn einmal die Zeitachse (T2) eingeführt ist, dann ist fraglich, ob jede

Versuchsperson diese auch wieder ausblenden kann, sollte sie erst danach mit T1 konfrontiert werden. Und noch kritischer erscheint uns die Aufgabe hinsichtlich einer umgekehrten Reihenfolge T3 / T2 zu sein: Wenn man den Versuchspersonen einmal die Stabilität der Quellen X und Y bekannt gegeben hat, wie garantiert man dann, dass sie diese Information danach ausblenden? Allein die Umbenennung in zwei völlig neue Quellen würde diesen Effekt nicht zwingend erzielen, da es ja nicht um die Abstraktion von den konkreten Quellen geht, sondern um die Tatsache, sich zuvor bereits in einer statistischen Umwelt mit expliziter Stabilitätsinformation bewegt zu haben und somit für die Frage der Stabilität sensitiviert worden zu sein. Diese Sensitivierung (so es einer solchen bedarf) ist, einmal geschehen, nicht wieder rückgängig zu machen.

Eine mögliche Lösung für dieses Problem bestünde darin, dass man auf abhängige Daten verzichtet und jede Versuchsperson nur eine der drei Bedingungen spielen lässt. Allerdings erscheint uns gerade der within-Vergleich sehr aussagekräftig zu sein, einmal weil er etwas über das Verständnis der Versuchspersonen hinsichtlich der Veränderung der Umwelten aussagt und zum anderen, weil er für T2 und T3 Aussagen darüber ermöglicht, ob One-Reason-Decision-Maker auf die Einführung der Zeitachse und der Stabilitätsinformation anders reagieren als More-Reason-Decision-Maker. Diese grundsätzlich unterschiedlichen Strategien sind in der bislang vorliegenden Literatur anhand repetitiver Einzeldurchgänge gut eingeführt (vgl. die in der Einleitung umfassend zitierten Arbeiten), so dass es sich anbietet, diese Personen anhand der Bedingung T1 zu identifizieren und dann zu schauen, wie weit sich die Veränderungen in der statistischen Umwelt auf die beiden grundsätzlichen Strategien auswirken.

Ohnehin erscheint es uns aussichtsreich, bei zukünftigen Experimenten in Design und Auswertung stärker auf diese Distinktion der Strategien (frugal versus absichernd) zu setzen. Auch im Trinkwasserszenario konnten diese Strategien als verhältnismässig stabil identifiziert werden (s. noch einmal Tabelle 10.3). Je komplexer die experimentell untersuchten Umwelten werden, umso variantenreicher wird auch die Verhaltensanpassung einer Person an die Veränderungen ausfallen. Und vor diesem Hintergrund scheint es ein guter Ausgangspunkt für die Auswertungen zu sein, wenn man etwas über die grundlegende Einstellung einer Person zur Stopppregel weiss.

Die Einführung der Zeitachse ist eine solche Erhöhung des Komplexitätsgrads. Die Versuchspersonen zeigen, dass sie damit umzugehen in der Lage sind, und zwar vornehmlich im Hinblick auf die Organisation ihres künftigen Verhaltens: Vor allem aus der zurückliegenden Test-Information und aus dem der Stabilitätsgrad der Wasserqualität der Quellen schliessen sie, bei welcher Alternative es sich lohnt weiterzumachen. Das macht ihr Verhalten auf Stichprobenebene effizienter (ohne dass diese Effekte allerdings so gross wären, dass sie bei jedem Individuum klar ablesbar wären. Dazu müssten vermutlich die Experimentaldurchgänge viel länger sein.) Bemerkenswert ist, dass sie dafür kein quantitatives Wissen über die Stabilität der Alternativen benötigen: Allein die Einführung der Zeitachse führt bereits zu den massgeblichen Verhaltensänderungen.

Nicht oder praktisch nicht genutzt wird die Stabilitätsinformation hingegen im Hinblick auf die Nutzung vergangener Information. Hier könnten die Versuchspersonen in einer wesentlich stärkeren Weise mutmassen, die ein Testresultat aus dem letzten Durchgang auch noch bei der aktuellen Wasserlieferung bestand hat, wenn die Stabilität so hoch ist wie bei Alternative Y (.90 nämlich). Dies geschieht jedoch nicht. Es wird deswegen für zukünftige Forschung spannend sein, eine grundsätzliche Asymmetrie der Informationsnutzung im Hinblick auf Vergangenheitsnutzung und Zukunftsplanung zu überprüfen und, sofern sie sich erhärten sollte, ihren psychologischen Gründen nachzuforschen.

10.6 Literatur

- Bröder, A. (2000a). Assessing the empirical validity of the "Take The Best"-heuristic as a model of human probabilistic inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 1332-1346.
- Bröder, A. (2000b). *"Take the best - Ignore the rest" Wann entscheiden Menschen begrenzt rational?* Lengerich: Pabst Science Publishers.
- Bröder, A. (2002). Take The Best, Dawes' Rule, and compensatory decision strategies: A regression-based classification method. *Quality & Quantity*, 36, 219-238.
- Bröder, A. (2003). Decision making with the "adaptive toolbox": Influence of environmental structure, intelligence, and working memory load. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 611-625.
- Bröder, A. (2005). *Entscheiden mit der "adaptiven Werkzeugkiste". Ein empirisches Forschungsprogramm.* Lengerich: Pabst Science Publishers.
- Bröder, A., & Schiffer, S. (2003). Take The Best versus simultaneous feature matching: Probabilistic inferences from memory and effects of representation format. *Journal of Experimental Psychology: General*, 132, 277-293.
- Bröder, A., & Schiffer, S. (2006). Adaptive flexibility and maladaptive routines in selecting fast and frugal decision strategies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(4), 904-918.
- Christen, S., Zurbruggen, S., Hausmann, D., & Läge, D. (2007). *Trinkwasser- und Börsenszenario: Aufbau, Ablauf und technische Spezifikation einer Experimentalumgebung zur entscheidungsvorbereitenden Informationssuche in zeitlichen Verläufen.* AKZ-Forschungsbericht 43. Zürich: Angewandte Kognitionspsychologie.
- Gigerenzer, G., & Goldstein, D. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, 103, 650-669.
- Gigerenzer, G., Todd, P. M., & Gerd Gigerenzer, A. R. (Eds.). (1999). *Simple Heuristics That Make Us Smart.* New York: Oxford University Press.
- Hausmann, D., Christen, S., & Läge, D. (2006). Urteils-Sicherheit um jeden Preis? Verletzung der Stoppregele in Abhängigkeit von der Höhe der Informationssuchkosten In B. Gula, R. Alexandrowicz, S. Strauss, E. Brunner, B. Jennull-Schiefer & O. Vitouch (Eds.), *Perspektiven psychologischer Forschung in Österreich. Proceedings zur 7. Wissenschaftlichen Tagung der Österreichischen Gesellschaft für Psychologie* (pp. 45-51). Lengerich: Pabst.
- Hausmann, D., & Läge, D. (2005). *Wenn Limiten zum Strategiewechsel führen: Wechsel der Suchstrategie unter situativen Einschränkungen.* AKZ-Forschungsbericht 06. Zürich: Angewandte Kognitionspsychologie.
- Hoffrage, U., Hertwig, R., & Czienskowski, U. (2003). *The ecological rationality of Take The Best's building blocks.* Vortrag an der 19th bi-annual conference on subjective Probability Utility and Decision Making (SPUDM). Zurich: Swiss Federal Institute of Technology.

- Kerstholt, J. H. (1994). The effect of time pressure on decision - making behavior in a dynamic task environment. *Acta Psychologica*, 86, 89-104.
- Läge, D., Hausmann, D., & Christen, S. (2005). *Wie viel bezahlen für eine valide Information? Suchkosten als limitierender Faktor der Informationssuche*. AKZ-Forschungsbericht 07. Zürich: Angewandte Kognitionspsychologie.
- Läge, D., Hausmann, D., Christen, S., & Daub, S. (2005). *Was macht einen "guten Cue" aus? Strategien der Informationssuche beim heuristischen Entscheiden unter Unsicherheit*. AKZ-Forschungsbericht 05. Zürich: Angewandte Kognitionspsychologie.
- Newell, B., Rakow, T., Weston, N., & Shanks, D. (2004). Search strategies in decision making: The success of 'success'. *Journal of Behavioral Decision Making*, 17, 117-137.
- Newell, B., & Shanks, D. (2003). Take the best or look at the rest ? Factors influencing one-reason decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 53-65.
- Newell, B., Weston, N., & Shanks, D. (2003). Empirical tests of a fast and frugal heuristic: not everyone "takes-the-best". *Organizational Behavior and Human Decision Processes*, 91, 82-96.
- Payne, J. W. (1976). Task complexity and contingent processing in decision making: An information search and protocol analysis. *Organizational Behavior and Human Performance*, 16, 366-387.
- Payne, J. W., & Bettman, J. R. (2000). Preferential Choice and Adaptive Strategy Use. In G. Gigerenzer & R. Selten (Eds.), *Bounded Rationality: The Adaptive Toolbox*. Cambridge, London: MIT Press.
- Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 534-552.
- Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. Cambridge: Cambridge University Press.
- Rieskamp, J., & Hoffrage, U. (1999). When do people use simple heuristics and how can we tell? In G. Gigerenzer, P. Todd & A. R. Group (Eds.), *Simple Heuristics That Make Us Smart* (pp. 141-168). New York: Oxford University Press.
- Rieskamp, J., & Hoffrage, U. (2006). *Inferences under time pressure: How opportunity costs affect strategy selection*. Submitted for publication.
- Zurbruggen, S., Christen, S., Hausmann, D., & Läge, D. (2008a). *Informationssuche beim Entscheiden in zeitlich dynamischen Umwelten*. AKZ-Forschungsbericht 44. Zürich: Angewandte Kognitionspsychologie.
- Zurbruggen, S., Christen, S., Hausmann, D., & Läge, D. (2008b). *Adaptivität der Informationssuche in statischen und in dynamischen Umwelten*. AKZ-Forschungsbericht 46. Zürich: Angewandte Kognitionspsychologie.
- Zurbruggen, S., & Läge, D. (2007). *Ressource Zeit statt Geld: Ein Vergleich zweier Szenarien zur Entscheidungsforschung*. AKZ-Forschungsbericht 45. Zürich: Angewandte Kognitionspsychologie.

Kapitel 11

Entscheiden in zeitlich dynamischen Umwelten: Suchen und Stoppen mit unterschiedlich stabilen Cues

Informationssuche als Grundlage einer Entscheidung wird in der Entscheidungsforschung üblicherweise durch wiederholte Einzeldurchgänge experimentell gemessen. Dabei sind diese Einzeldurchgänge als unabhängig zu betrachten, d.h. Probanden können keine Informationen aus vorhergehenden Durchgängen für die aktuelle Entscheidung nutzen noch können sie zum aktuellen Zeitpunkt bereits Informationen für zukünftige Entscheidungen suchen. Die Einführung einer Zeitachse und eines Grades an Cue-Stabilität im Rahmen eines Börsenszenarios ermöglicht nun den Einbezug von Vergangenheit und Zukunft. Es zeigt sich, dass Versuchspersonen in einer zeitlich dynamischen Umwelt mit weniger Information auszukommen glauben, dass sie teilweise ganz ohne neue Informationssuche entscheiden und dass sie Cues mit hoher Stabilität (die also für zukünftige Entscheidungen noch verwendbar sind) in stärkerem Masse berücksichtigen als in einer Vergleichsbedingung ohne Zeitachse. In einer bedeutenden Zahl von Fällen ziehen sie die längerfristig stabile Information bei der Informationssuche sogar der momentan valideren Information vor. Interessant ist, dass die beschriebenen Verhaltensänderungen sich vornehmlich auf ein Drittel der Personen beziehen, und zwar auf solche, die in einer Bedingung ohne zeitliche Dynamik One Reason Decision Making vorziehen.

11.1 Einleitung

11.1.1 Entscheiden im Zeitverlauf

Entscheidungen brauchen Information. Diese Maxime, welche das Verhalten von Menschen in ihrer Umwelt besser macht als der Zufall, dürfte in der Forschung unbestritten sein. Und genauso unbestritten dürfte es sein, dass Menschen in der Lage sein müssen, a) sich diese Informationen zu beschaffen und b) sie so auszuwerten, dass sie auf dieser Basis die entsprechenden angemessenen Entscheidungen treffen können. Letzteres ist der traditionelle Fokus der Entscheidungspsychologie, nämlich diejenigen kognitiven Verfahren ausfindig zu machen, mit denen vorliegende Informationen in eine Entscheidung transformiert werden. In der experimentellen Forschung reicht es dazu aus, Personen mit einem (im Sinne der jeweiligen Fragestellung möglichst geschickt kombinierten) Satz von Informationen zu konfrontieren und zu beobachten, welche Entscheidungen sie treffen. Seit der Entwicklung des Information Board (z.B. in Form des *Mouselab* von Payne, 1976) und vor allem seit dem Vorschlag sequentiell arbeitender Urteilsheuristiken (z.B. Gigerenzer & Goldstein, 1996) ist ausserdem auch die Frage der Informationsbeschaffung selbst immer stärker in den Fokus der Aufmerksamkeit von Entscheidungsforschern gerückt. Sequentielle Urteilsheuristiken, von denen *Take The Best* vermutlich die prominenteste ist, kombinieren eben Informationssuche und Entscheidungsfindung, indem sie Suchregel, Stoppregel und Entscheidungsregel als aufeinander bezogene Bausteine eines Gesamtverhaltens verstehen. Wie stark die Richtung einer Entscheidung von der vorausgehenden Informationssuche abhängt, konnten Lägre & Hausmann (2007) zeigen, indem sie beim Einsatz eines *Information Board* für den Mehralternativenfall fanden, dass sich die Entscheidungen einer Person in derselben Situation, nur einmal mit vorgegebener und einmal mit selbst gesuchter Information, so radikal voneinander unterscheiden als würde man einen Probanden gegen einen beliebig anderen des untersuchten Samples austauschen.

Die Tradition der Forschung mit dem Information Board besteht, gleich ob mit aktiver Informationssuche oder mit vorgegebener Information, darin, der Versuchsperson eine Reihe von Durchgängen zu präsentieren, welche jeweils eine bestimmte Umwelt repräsentieren. In dieser Umwelt haben die Personen dann Entscheidungen zu treffen, und dazu können sie die angebotenen Informationen nutzen. Pro Aufgabe, Informationen und Kontextvariablen lassen sich so Erkenntnisse über den Entscheidungsprozess gewinnen. Die einzelnen Durchgänge (in der Regel einige Dutzend bis einige hundert) sind zumeist als punktuelle Messwiederholungen einer begrenzten Zahl von Variationen aufgebaut. Diese Messwiederholungen ermöglichen später eine statistische Auswertung, um die Strategien probabilistisch gegeneinander und/oder gegen den Zufall zu testen.

Diese Form der punktuellen Messwiederholung entspricht einem bewährten Standard in der experimentellen Psychologie. Was sie nicht berücksichtigen kann (und auch bei den meisten Fragestellungen nicht zu berücksichtigen braucht), ist der Umstand, dass sich Entscheidungsverhalten nicht zu isolierten Zeitpunkten abspielt, sondern in der Realität immer in einen Zeitfluss eingebettet ist. Insofern ist die grundsätzliche Idee des *Information Board* als experimentalpsychologi-

schες Verfahren eine Abstraktion der Wirklichkeit. Gleichwohl ist es aber auch möglich, diese Abstraktion auszuweiten und den natürlichen Zeitfluss mit einzubeziehen. Das im Folgenden zu berichtende Experiment wird das tun und eine Zeitachse einführen, auf der einzeldurchgangsübergreifend Informationen gesucht und für die zu treffenden Entscheidungen verwendet werden können.

Zwar haben verschiedene Entscheidungsforscher die Zeitdimension bereits in ihre Untersuchungen mit einfließen lassen (z.B. Betsch, Haberstroh, Glöckner, Haar, & Fiedler, 2001, in Bezug auf routiniertes Verhalten; Busemeyer & Townsend, 1993, in Bezug auf Präferenzen; Kerstholt & Raaijmakers, 1997, in Bezug auf den richtigen Zeitpunkt einer Entscheidung), aber es fehlt bislang an Experimenten zum Informationssuchverhalten in Umwelten, die sich über die Zeit verändern, und zum Entscheidungsverhalten in einer Welt von sich ständig ändernden Informationen. Diese Umwelten können wir auch als „dynamisch“ bezeichnen, um sie von den „statischen“ Umwelten abzugrenzen, in denen normalerweise das *Information Board* mit den punktuellen Messwiederholungen Entscheidungsverhalten erhebt.

Die Einführung einer Zeitachse bedeutet für das Entscheiden auf der Basis probabilistischer Informationen zweierlei: Zum einen können Informationen aus der Vergangenheit für die aktuelle Entscheidung genutzt werden. Dazu muss man allerdings zumindest eine Ahnung haben, wie stabil denn die Zielvariablen sind (also die Alternativen, für oder gegen die man sich entscheidet oder aus denen man eine auswählt). Sind sie sehr variabel, so verliert zeitlich zurückliegende Information schnell an Validität. Neben diesem möglichen Blick in die Vergangenheit kann man aber auch in die Zukunft hinein planen und bereits jetzt nach Informationen suchen, die für eine zukünftige Entscheidung oder Wahl relevant werden können. Um das mit Erfolg zu tun, braucht man zusätzliches Wissen über die Stabilität der Prädiktorvariablen, also der probabilistischen Cues. Wenn sich ein Prädiktor über die Zeit als sehr stabil erweist, dann kann es sich lohnen, diese Information beizeiten zu besorgen und über mehrere Entscheidungen hin zu verwenden; ist er hingegen sehr variabel, so ist seine Information nur für den Moment geeignet.

Beide hiermit neu eingeführten Werte – die Stabilität der Alternativen sowie die Stabilität der Cues – hängen über die Validität der Cues in einem Punkt miteinander zusammen: Ein Cue kann nicht zugleich hochgradig valide sein, sich aber in der Stabilität stark von der der Alternativen unterscheiden. Hohe Validität setzt also ähnliche Stabilität von Prädiktor und Zielvariablen voraus. Im mittleren und unteren Validitätsbereich hingegen gilt diese Einschränkung hinsichtlich der Stabilitäten nicht mehr.

Das im Folgenden berichtete Experiment greift auf das in der Entscheidungspsychologie vertraute Börsenszenario zurück (Bröder, 2000b; Bröder & Schiffer, 2006; Läge & Hausmann, 2007; Newell, Rakow, Weston, & Shanks, 2004; Newell & Shanks, 2003; Newell, Weston, & Shanks, 2003; Rakow, Newell, Fayers, & Hersby, 2005), variiert dieses allerdings in drei Komponenten erheblich. Im Kern steht die Einführung der Zeitachse sowie des Stabilitätsfaktors als Attribut a) der Zielvariablen und b) der Prädiktorvariablen. Das Wissen einer Versuchsperson um die Stabilität der Zielvariablen ermöglicht ihr den Blick in die Vergangenheit. Wie dabei aus unterschied-

lichen Stabilitäten der Zielvariablen aus vergangenen Ausprägungen auf die derzeitige Ausprägung geschlossen wird, wurde in einem anderen Szenario bereits untersucht (zu diesem Trinkwasserszenario siehe Zurbriggen, Hausmann, Christen, & Läge, 2007). Das Wissen um die Stabilität der Prädiktorvariablen gibt der Versuchsperson zusätzlich die Möglichkeit, aus der gegenwärtigen Ausprägung dieser Prädiktorvariablen zukünftige Entscheidungen vorzubereiten.

11.1.2 Such- und Stoppregel in zeitlich dynamischen Umwelten

Im Rahmen der Forschung über schrittweise Urteilsheuristiken sind Such-, Stopp- und Entscheidungsregel als Bausteine einer Heuristik zu unterscheiden (so wie am Beispiel der *Take The Best*-Heuristik von Gigerenzer & Goldstein, 1996, eingeführt und in zahlreichen Folgepublikationen stärker herausgearbeitet; wird ab hier mit TTB abgekürzt). Dabei haben sich vor allem die Suchregel und die Stoppregel als für die Forschung fruchtbar erwiesen.

Bei der *Suchregel* geht es vor allem um die Frage, welche Qualitäten einen probabilistischen Cue attraktiv machen (und damit die Reihenfolge festlegen, mit der Cues in der Informationssuche befragt werden). Der ursprüngliche Vorschlag war die Validität (Gigerenzer & Goldstein, 1996; Gigerenzer, Hoffrage, & Kleinbölting, 1991), was verknüpft mit dem *One Reason Decision Making* zur *Take The Best*-Heuristik und bei der Integration mehrerer Cue-Informationen zu einem gewichteten Modell führt (vgl. auch Bröder, 2000a; Bröder, 2000b). In dem Moment, in dem Cues allerdings nicht immer diskriminieren, wird bei einer ressourcenaufwändigen Informationssuche auch *Success* oder *Usefulness* (also Kombinationen von Validität und Diskriminationsrate) bzw. die Diskriminationsrate selbst zu einer valablen Alternative für die Bestimmung der Qualitätsrangfolge von Cues (vgl. Läge, Hausmann, & Christen, in prep; Newell et al., 2004). Durch die Einführung einer Stabilitätsrate für die Cues kommt nun eine weitere Cue-Qualität ins Spiel: Ein besonders stabiler, gleichzeitig aber auch vernünftig valider Cue wird attraktiver, und zwar möglicherweise sogar attraktiver als ein deutlich validerer, aber instabiler Cue. Das könnte eine rein an der Validität orientierte Suchregel (so wie sie anzunehmen ist, wenn nichts als die Validität als Gütekriterium zur Verfügung steht) somit substantiell verändern. Voraussetzung dafür (und gleichzeitig Erklärungsmuster für ein entsprechendes Verhalten) ist die Intention einer Person, Cues bereits mit Blick auch auf zukünftige Entscheidungssituationen zu befragen.

Bei der *Stoppregel* geht es in erster Linie um die Unterscheidung zwischen *One Reason Decision Making* (stoppen nach einem diskriminierenden Cue) und *More Reason Decision Making* (fortsetzen der Suche, auch wenn bereits ein Cue diskriminierte, um eine breitere Informationsgrundlage für die Entscheidung zu besitzen). Diese beiden Formen des Entscheidens kreieren also ganz unterschiedliche Stoppregeln, und wie es scheint, haben Individuen verhältnismässig stabile Präferenzen für das Entscheiden auf die eine oder die andere Weise (Bröder, 2000a, 2000b; Zurbriggen, Christen, Hausmann, & Läge, 2008; Zurbriggen & Läge, 2007). Allerdings finden sich starke Hinweise dafür, dass eine bestimmte (interindividuell womöglich sehr stark schwankende) Schwelle an Urteilssicherheit die Stoppregel moderiert: Sind die Validitäten der besten diskriminierenden Cues eher niedrig, tendieren die Personen zur Fortsetzung der Informationssuche, sind die Validitäten hingegen sehr hoch, dann stoppen sie gerne nach einer einzigen diskriminierenden Information (Hausmann & Läge, 2005).

Vor dem Hintergrund des *One Reason Decision Making* gewinnt die Einführung einer Zeitachse an Brisanz, denn damit liegen womöglich bereits Informationen aus vergangenen Entscheidungsdurchgängen vor und können für die aktuelle Entscheidung genutzt werden. Werden diese Informationen als in ihrer Validität noch hinreichend hoch erlebt, so müsste sich – zumindest bei Personen, die unter den entsprechenden Umständen dem *One Reason Decision Making* zugeeignet sind – eine Tendenz finden lassen, ab und zu ganz ohne neue Informationssuche zu entscheiden. Wie Hausmann, Christen & Läge (2006) in einem Experiment mit variierenden Informationskosten zeigen konnten, vermeiden es Personen normalerweise strikt, ohne jegliche Information zu entscheiden. Wenn jemand aufgrund alter Informationen eine Entscheidung trifft, ohne zuvor eine neue Suche gestartet zu haben, so werden wir dies in der Auswertung der Daten als *No New Reason Stopping Rule* (NNRStR) bezeichnen, um es der nach einem diskriminierenden Cue gestoppten Informationssuche (*One Reason Stopping Rule*, ORStR) und der auch nach einem diskriminierenden Cue noch fortgesetzten Informationssuche (*More Reason Stopping Rule*, MRStR) als dritte Stoppregel zur Seite zu stellen.

11.1.3 Hypothesen

Aus dem Gesagten lassen sich folgende Hypothesen ableiten. Zunächst einmal sollte die Einführung einer Zeitachse dazu führen, dass bei halbwegs stabilen Zielvariablen die Information aus der Vergangenheit für die aktuelle Entscheidung mit hinzugezogen wird. Messbare Folge ist erstens, dass die Zahl gesuchter Informationen sinkt (Hypothese 1). Zweitens sollten Personen, die sich unter den experimentellen Bedingungen der „traditionellen“ Forschung (also ohne Zeitachse) als *One Reason Decision Maker* klassifizieren lassen, eine Tendenz zeigen, zumindest in geeigneten Fällen (ausreichend hohe verbleibende Validität der Information aus der Vergangenheit) vor einer aktuellen Entscheidung ganz auf Informationssuche zu verzichten (Hypothese 2). Und drittens sollte mit dem Wissen um den Stabilitätskoeffizienten von Cues ein Cue mit besonders hoher Stabilität an Attraktivität gewinnen, d.h. häufiger gewählt werden als ohne Wissen um die Stabilität (Hypothese 3a) und – je nach Wahrnehmung der Gesamtattraktivität durch die individuelle Versuchsperson – auch häufig als erster Cue konsultiert werden als ein eigentlich deutlich validerer, aber zeitlich wenig stabiler Cue (Hypothese 3b).

11.2 Methode

11.2.1 Aufbau des Experiments

Das Experiment besteht aus drei sequentiellen Bedingungen mit je 20 Durchgängen. Die erste Bedingung ist die statische Vergleichsbedingung. Hier erleben die Probanden die Ereignisse als voneinander unabhängige Durchgänge (so wie in bisherigen Entscheidungsexperimenten mit *Information Board*-Designs). Die Versuchspersonen können in jeder der 20 Runden („Handelstage“) in einen von zwei Optionsscheinen investieren, wobei jeweils zwei neue Optionsscheine zur Auswahl stehen. Das Information Board besteht aus einer 2x4-Matrix (zwei Alternativen x vier Cues) mit der zusätzlichen Möglichkeit, den investierbaren Betrag nicht auf eine der beiden Alternativen zu setzen, sondern als Sparbetrag direkt dem Gewinnkonto gutzuschreiben. Offen ersichtlich sind die Validitäten der Cues und die Schaltflächen für den Informationsabruf

(siehe Abbildung 11.1). Die Validitäten der Cues werden in ganzen Zahlen angegeben (als Treffer über 100 Durchgänge eingeführt): 90, 85, 80 und 75. Die Reihenfolge der Cues ist bei jeder Person zufällig, bleibt dann aber über die drei Bedingungen des Experiments konstant.

Informationen	Trefferrate	Optionsschein X	Optionsschein Y
Indikator A	80	?	?
Indikator B	75	?	?
Indikator C	90	↓	↓
Indikator D	85	?	?

Handel						
Budget 10'000 minus Suchkosten 2'000 = investierbarer Betrag 8'000 <div style="display: flex; align-items: center;"> <div style="margin-right: 10px;"> ? ? ? </div> <div> <table border="1"> <tr> <td>Optionsschein X</td> <td>0</td> </tr> <tr> <td>Optionsschein Y</td> <td>0</td> </tr> <tr> <td>Sparbüchse</td> <td>8'000</td> </tr> </table> </div> <div style="margin-left: 10px;"> → → → </div> <div style="border: 1px solid black; padding: 10px; text-align: center;"> Auswahl bestätigen </div> </div>	Optionsschein X	0	Optionsschein Y	0	Sparbüchse	8'000
Optionsschein X	0					
Optionsschein Y	0					
Sparbüchse	8'000					

Übersicht
Bisheriger Kontostand : 275'000 Gewinn in den letzten 5 Tagen : 89'000 (darunter 1 Tage ohne Gewinn) Neuer Kontostand : 364'000 Anzahl vergangener Handelstage : 19

Weiter

Abbildung 11.1: Information Board der 1. Börsenbedingung

Den Versuchspersonen steht pro Runde ein Budget von 10'000 Franken zur Verfügung, welches sie für Investitionen und für die Informationssuche einsetzen können. Investieren sie einen Betrag in einen Optionsschein, so verdreifacht sich der Wert im Fall der positiven Entwicklung dieser Anlage, im Fall der negativen Entwicklung verfällt der Betrag.

Die Probanden sind darüber informiert, dass die Wahrscheinlichkeit, dass sich ein Optionsschein positiv entwickelt, jeweils 33% beträgt und dass zwei Drittel der Optionsscheine eine negative Entwicklung haben werden. Die Versuchspersonen können jeweils in genau einen Optionsschein investieren oder den Betrag direkt auf das Gewinnkonto transferieren. In diesem Fall erhalten sie genau diesen Betrag gutgeschrieben, ohne irgendetwas zu riskieren.

Um die Entscheidung auf sicherere Füße zu stellen, stehen die vier Cues zur Verfügung. Diese können unabhängig voneinander betrachtet werden, und zwar Alternative für Alternative. (Diese Unterscheidung ist wichtig: Die Probanden betrachten nicht einen Cue mit seiner Information für alle Alternativen, sondern jeweils nur für eine Alternative. Auf diese Weise wird die Diskriminationsrate als zusätzlicher Faktor aus dem Design herausgehalten). Die Information besteht entweder aus einem roten, nach unten zeigenden Pfeil (Vorhersage, dass der Optionsschein seinen

Wert verlieren wird) oder aus einem grünen, nach oben zeigenden Pfeil (Vorhersage, dass der Optionsschein an Wert gewinnen wird). Die angegebenen Validitäten zu jedem Cue sind in Bezug auf jede Bedingung korrekt, so dass sich die Probanden mit der entsprechenden Wahrscheinlichkeit auf diese Vorhersage verlassen können.

Die Kosten für das Anschauen einer Cue-Information betragen 1'000 Franken. Der Betrag für angeschaute Cues wird unmittelbar dem Kundenkonto abgezogen und kann nicht mehr investiert werden. Nach jeweils drei bis fünf Runden – das exakte Intervall wird jedes Mal neu durch Zufall bestimmt – sowie in der letzten Runde einer jeden Bedingung erhalten die Versuchspersonen Feedback über die im letzten Intervall erzielte Leistung, und zwar hinsichtlich der Höhe des im Intervall erzielten Gewinns sowie des bisher erreichten Gesamtgewinns. Ausserdem werden sie über die Dauer des abgelaufenen Intervalls und die Anzahl Tage (im Intervall) ohne Gewinn informiert.

Informationen	Trefferrate	Fonds X mit Stabilität = 80	Fonds Y mit Stabilität = 80
Indikator A	80	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
Indikator B	75	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
Indikator C	90	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
Indikator D	85	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
Vergangene Tage seit Erhalt der Info :		4 3 2 1 0	4 3 2 1 0

Handel
Budget 10'000 minus Suchkosten 1'000 = investierbarer Betrag 9'000 <div style="display: flex; align-items: center;"> <div style="margin-right: 10px;"> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> </div> <div> <div style="border: 1px solid black; padding: 5px; margin-bottom: 5px;">Fonds X 9'000</div> <div style="border: 1px solid black; padding: 5px; margin-bottom: 5px;">Fonds Y 0</div> <div style="border: 1px solid black; padding: 5px; margin-bottom: 5px;">Sparbüchse 0</div> </div> <div style="margin-left: 20px;"> <div style="border: 1px solid black; padding: 10px; text-align: center;">Auswahl bestätigen</div> </div> </div>

Übersicht
Bisheriger Kontostand : 184'000 Gewinn in den letzten ? Tagen : ? (darunter ? Tage ohne Gewinn) Neuer Kontostand : ? Anzahl vergangener Handelstage : 14

Nächster Handelstag

Abbildung 11.2: Information Board der 2. Börsenbedingung

In der zweiten Bedingung tauchen die Personen nun in einen Zeitfluss ein (s. Abbildung 11.2). Neben der Einführung der Stabilität der Alternativen bleiben die angeschauten Cue-Informationen der letzten vier Runden jeweils auf dem Information Board sichtbar. Die beiden Alternativen werden nicht mehr als Optionsscheine, sondern als Fonds bezeichnet – dies aus Plausibilitätsgründen, da nun immer das gleiche Alternativenpaar zur Auswahl steht. Den Versuchspersonen

sonen ist erklärt, dass diese beiden Fonds eine Stabilität von 80 haben, also an vier von fünf Handelstagen dieselbe Ausprägung (Wertzuwachs oder Wertverlust) besitzen wie am Vortrag.

Informationen	Trefferrate	Stabilität	Fonds X mit S = 80					Fonds Y mit S = 80				
Indikator A	80	95	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Indikator B	75	95	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Indikator C	90	65	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Indikator D	85	65	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Anzahl vergangene Tage seit Erhalt der Information			4	3	2	1	0	4	3	2	1	0

Handel
Budget 10'000 minus Suchkosten 1'000 = investierbarer Betrag 9'000 <div style="display: flex; align-items: center;"> <div style="margin-right: 10px;"> <input type="checkbox"/> Fonds X <input type="checkbox"/> Fonds Y <input type="checkbox"/> Sparbüchse </div> <div> <div style="border: 1px solid black; padding: 5px; margin-bottom: 5px;"> 9'000 </div> <div style="border: 1px solid black; padding: 5px; margin-bottom: 5px;"> 0 </div> <div style="border: 1px solid black; padding: 5px;"> 0 </div> </div> </div> <div style="border: 1px solid black; padding: 10px; margin-top: 10px; text-align: center;"> Auswahl bestätigen </div>

Übersicht
<div style="border: 2px solid red; padding: 5px;"> Bisheriger Kontostand : 0 Gewinn in den letzten 5 Tagen : 84'000 (darunter 2 Tage ohne Gewinn) Neuer Kontostand : 84'000 </div> <div style="margin-top: 10px;"> Anzahl vergangener Handelstage : 4 </div> <div style="border: 1px solid black; padding: 10px; margin-top: 10px; text-align: center;"> Nächster Handelstag </div>

Abbildung 11.3: Information Board der 3. Börsenbedingung

Die dritte Bedingung unterscheidet sich von der zweiten Bedingung lediglich darin, dass zusätzlich die Stabilitäten der Cues explizit auf dem Information Board angegeben werden (Abbildung 11.3). Die Stabilität der beiden valideren Cues ($V = .90$ bzw. $V = .85$) beträgt .65, die Stabilität der beiden weniger validen Cues ($V = .80$ bzw. $V = .75$) beträgt hingegen .95. Die Stabilitäten der Tests sind somit gegenläufig zu den Validitäten, so dass die weniger validen Cues dann attraktiver werden sollten, wenn jemand sie sich auch schon mit Blick auf künftige Handelstage anschauen möchte. Die Stabilitätsraten der Cues werden im Experiment als ganze Zahlen angegeben.

Eine detaillierte technische Beschreibung der Experimentalumgebung sowie der verwendeten Verläufe von Prädiktor- und Zielvariablen findet sich bei Christen, Zurbriggen, Hausmann & Läge (2007).

11.2.2 Versuchsdurchführung

Das Experiment ist in Microsoft Visual Basic 6.0 programmiert und wurde vollständig auf einem PC mit Windows XP-Betriebssystem durchgeführt. Alle Personen absolvierten die drei Versuchsbedingungen des Börsenszenarios in derselben Reihenfolge. (Da sich ja jeweils die Infor-

mation über die Cue-Güte erweitert, ist es nicht sinnvoll, die Reihenfolge zu variieren. Sonst würden die Versuchspersonen womöglich ihr Wissen aus einer vorausgehenden Bedingung mit in die nächste Bedingung nehmen.)

Das Experiment wurde in Verbindung mit einem anderen Entscheidungsexperiment (in einem Trinkwasserszenario) durchgeführt. Dort waren zunächst drei Durchgänge (Versuchspersonengruppe 1) bzw. zwei Durchgänge (Versuchspersonengruppe 2) zu absolvieren, bevor die drei Bedingungen des Börsenszenarios gespielt wurden.

11.2.3 Stichprobe

Die Stichprobe bestand aus 60 Versuchspersonen (30 Männern und 30 Frauen), darunter 38 Psychologie-Studierende der Universität Zürich. Das Durchschnittsalter betrug 30.3 Jahre (range 19–60, SD = 8.9). Die Psychologie-Studierenden müssen im Rahmen ihrer Ausbildung an der Universität Zürich mindestens fünf Stunden an Experimenten teilnehmen, an welchen ist ihnen dabei frei gestellt. Die Nicht-Studierenden nahmen freiwillig teil. Unter dem Drittel mit der besten Performance wurde eine Gewinnsumme von zweimal 100 Franken verlost.

11.3 Resultate

Um den Einfluss der dynamischen Umwelt auf das Such- und Stoppverhalten zu analysieren, wurde bei den im Folgenden berichteten inferenzstatistischen Auswertungen (t-Tests für abhängige Stichproben) $\alpha = 0.05$ mit zweiseitiger Hypothesenprüfung als Grundlage für das Akzeptieren eines Mittelwertsunterschiedes gewählt.

11.3.1 Kaufhäufigkeiten aller Cues

In der ersten Bedingung wurden in den 20 Runden im Mittel 40.6 Cues (SD=14.1) gekauft, in der zweiten Bedingung 36.9 Cues (SD=16.7) und in der dritten Bedingung 37.0 Cues (SD=17.3). Die Abnahme der Kaufhäufigkeit bei der Einführung eines zeitlichen Kontextes in Bedingung 2 gegenüber der Standardbedingung 1 ist signifikant ($t = 3.01$; $p < .01$). Dabei korrelieren die Häufigkeiten der beiden Bedingungen über die Probanden mit $r = .82$. Zwischen der zweiten und dritten Bedingung unterscheidet sich die Kaufhäufigkeiten aller Cues nicht signifikant ($t = 0.14$; $p = .89$), die Korrelation über die Probanden bleibt mit $r = .87$ hoch.

Die Versuchspersonen reagieren mit der Gesamthäufigkeit also auf die Einführung der Zeitachse (und die damit verbundene Möglichkeit, Informationen aus der Vergangenheit benutzen zu können). Die hohe Korrelation macht deutlich, dass das Gesamtmuster der Kaufhäufigkeiten über die Probanden in etwa erhalten bleibt.

11.3.2 Kaufhäufigkeiten einzelner Cues

Mit der zusätzlichen Einführung der Cue-Stabilitäten in Bedingung 3 ändert sich an der Gesamthäufigkeit der Cuekäufe nichts. Wie Tabelle 11.1 aber zeigt, reagieren die Versuchspersonen trotzdem auf die neue Situation, indem sie nämlich die beiden weniger validen, aber zeitlich sehr stabilen Cues massiv stärker berücksichtigen: Der Kauf von Cue 3 verdreifacht sich (von Bedin-

gung 2 mit $M_2 = 3.6$ auf $M_3 = 10.9$ in Bedingung 3; $t = 5.19$; $p < .01$) und der Kauf von Cue 4 verdoppelt sich (von $M_2 = 1.8$ auf $M_3 = 3.6$; $t = -2.57$; $p < .05$). Dies geht vor allem zu Lasten von Cue 1, der in Bedingung 2 mit $M_2 = 24.8$ wesentlich häufiger gekauft wurde als in Bedingung 3 ($M_3 = 17.1$; $t = 5.23$; $p < .01$). Der Rückgang von Cue 2 von $M_2 = 6.7$ auf $M_3 = 5.5$ erweist sich als nicht signifikant ($t = 1.43$; $p = .16$).

Tabelle 11.1: Summe der einzelnen Kaufhäufigkeiten in den drei Bedingungen B1, B2 und B3 (je 20 Runden); Mittelwerte und Standardabweichungen pro Versuchsperson über alle Runden.

	B1	B2	B3
Cue 1	1718	1489	1024
($N = .90$)	$M_1 = 28.6$	$M_2 = 24.8$	$M_3 = 17.1$
($S = .65$)	$SD_1 = 8.3$	$SD_2 = 10.3$	$SD_3 = 12.1$
Cue 2	380	402	330
($N = .85$)	$M_1 = 6.3$	$M_2 = 6.7$	$M_3 = 5.5$
($S = .65$)	$SD_1 = 7.2$	$SD_2 = 7.9$	$SD_3 = 7.7$
Cue 3	192	214	653
($N = .80$)	$M_1 = 3.2$	$M_2 = 3.6$	$M_3 = 10.9$
($S = .95$)	$SD_1 = 4.6$	$SD_2 = 5.1$	$SD_3 = 10.4$
Cue 4	143	107	214
($N = .75$)	$M_1 = 2.4$	$M_2 = 1.8$	$M_3 = 3.6$
($S = .95$)	$SD_1 = 4.4$	$SD_2 = 3.7$	$SD_3 = 5.7$

Bemerkenswert sind die hohen STD in Bedingung 3 bei Cue 1 und 3. Dies deutet auf (mindestens) zwei grundsätzlich verschiedene Muster hin, mit denen die Probanden auf die Einführung der Cue-Stabilität reagieren. Wir werden weiter unten darauf zurückkommen und diese Gruppen zu identifizieren.

Ein weiteres Indiz für mögliche Subgruppen sind die durchgängig niedrigeren Korrelationen der Kaufhäufigkeiten für die einzelnen Cues im Vergleich Bedingung 3 / 2 gegenüber Bedingung 2 / 1: Zwischen Bedingung 2 und 1 korrelieren die Cues folgendermassen: Cue 1: $r = .78$, Cue 2: $r = .82$, Cue 3: $r = .61$, Cue 4: $r = .57$. Zwischen Bedingung 3 und 2 sind alle vier Werte deutlich niedriger: Cue 1: $r = .49$, Cue 2: $r = .66$, Cue 3: $r = .15$, Cue 4: $r = .41$.

Nachzutragen bleibt noch der Häufigkeitsvergleich zwischen Bedingung 1 und Bedingung 2: Hier unterscheidet sich einzig die Kaufhäufigkeit beim validesten Cue signifikant (Cue 1: $t = 4.54$; $p < .01$). Bei den weniger validen Cues hingegen gibt es keine signifikanten Unterschiede in der Kaufhäufigkeit (Cue 2: $t = 0.62$; $p = .54$; Cue 3: $t = 0.66$; $p = .52$; Cue 4: $t = 1.22$; $p = .23$). Der weiter oben berichtete signifikante Haupteffekt für die Gesamthäufigkeit zwischen Bedingung 1 und Bedingung 2 geht also voll zu Lasten des validesten Cues! In der Konsequenz muss dies bedeuten, dass die Versuchspersonen in Bedingung 2 häufiger einmal ganz ohne Cue-Kauf auskommen, und zwar primär in solchen Fällen, in denen sie in Bedingung 1 lediglich den Cue 1 angeschaut hatten. Da gleichzeitig mit dem sinkenden Mittelwert die Standardabweichung steigt (von $SD_1 = 8.3$ auf $SD_2 = 10.3$), deutet sich auch hier die Existenz (mindestens) zweier Subgruppen mit unterschiedlichem Verhaltensmuster an.

Tabelle 11.2: Prozentuale Suche bei den einzelnen Cues (1-4) in den 3 Bedingungen, sortiert nach Cue1. In den zwei letzten Spalten sind die Veränderungen von Bedingung 1 zu 2 bzw. von 2 zu 3 kategorisiert.

Vp-ID	1. Bedingung				2. Bedingung				3. Bedingung				Veränderung zw. Bedingungen	
	1	2	3	4	1	2	3	4	1	2	3	4	1 vs. 2	2 vs. 3
14	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	NEIN	NEIN
31	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	NEIN	NEIN
32	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	NEIN	NEIN
39	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	NEIN	NEIN
41	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	NEIN	NEIN
46	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	NEIN	NEIN
47	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	NEIN	NEIN
52	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	NEIN	NEIN
53	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	NEIN	NEIN
54	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	NEIN	NEIN
58	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	NEIN	NEIN
5	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.9	0.0	0.1	0.0	NEIN	NEIN
23	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.9	0.0	0.1	0.0	NEIN	NEIN
10	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.5	0.0	0.5	0.0	NEIN	JA
25	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.5	0.0	0.5	0.0	NEIN	JA
13	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.4	0.5	0.1	0.0	NEIN	NEIN
22	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.3	0.0	0.7	0.0	NEIN	JA
8	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	NEIN	JA
33	1.0	0.0	0.0	0.0	0.4	0.4	0.2	0.1	0.4	0.4	0.2	0.1	JA	NEIN
43	1.0	0.0	0.0	0.0	0.1	0.0	0.9	0.0	0.1	0.0	0.9	0.0	JA	NEIN
6	0.9	0.0	0.1	0.0	0.9	0.0	0.1	0.0	0.7	0.1	0.1	0.1	NEIN	NEIN
4	0.9	0.1	0.0	0.0	0.8	0.2	0.0	0.0	0.7	0.1	0.2	0.0	NEIN	NEIN
12	0.9	0.1	0.0	0.0	0.8	0.2	0.0	0.0	0.6	0.0	0.4	0.0	NEIN	JA
17	0.9	0.1	0.0	0.0	0.7	0.2	0.1	0.0	0.7	0.2	0.1	0.0	NEIN	NEIN
30	0.9	0.0	0.1	0.0	0.5	0.1	0.4	0.0	0.4	0.1	0.4	0.0	JA	NEIN
56	0.8	0.1	0.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	NEIN	NEIN
24	0.8	0.2	0.0	0.0	0.8	0.1	0.1	0.1	0.9	0.1	0.0	0.0	NEIN	NEIN
40	0.8	0.0	0.0	0.1	0.8	0.1	0.0	0.1	0.8	0.1	0.0	0.1	NEIN	NEIN
29	0.8	0.2	0.0	0.0	0.8	0.2	0.0	0.0	0.3	0.0	0.6	0.0	NEIN	JA
18	0.7	0.2	0.0	0.1	0.9	0.0	0.0	0.0	0.5	0.0	0.5	0.0	NEIN	JA
3	0.7	0.2	0.0	0.1	0.9	0.1	0.0	0.0	0.1	0.0	0.7	0.2	NEIN	JA
2	0.7	0.0	0.3	0.0	0.8	0.0	0.2	0.0	0.7	0.0	0.2	0.1	NEIN	NEIN
9	0.7	0.2	0.1	0.0	0.7	0.3	0.0	0.0	0.8	0.1	0.2	0.0	NEIN	NEIN
1	0.7	0.3	0.0	0.0	0.7	0.3	0.0	0.0	0.7	0.3	0.0	0.0	NEIN	NEIN
21	0.7	0.3	0.0	0.1	0.7	0.2	0.0	0.0	0.7	0.2	0.0	0.0	NEIN	NEIN
44	0.7	0.3	0.1	0.0	0.6	0.3	0.1	0.0	0.6	0.3	0.1	0.0	NEIN	NEIN
51	0.7	0.2	0.1	0.1	0.5	0.2	0.2	0.1	0.5	0.2	0.2	0.1	NEIN	NEIN
19	0.7	0.3	0.1	0.0	0.5	0.3	0.2	0.0	0.4	0.4	0.2	0.1	NEIN	NEIN
36	0.7	0.0	0.3	0.0	0.3	0.2	0.2	0.2	0.3	0.2	0.2	0.2	JA	NEIN
42	0.6	0.3	0.1	0.1	0.8	0.1	0.0	0.0	0.8	0.1	0.0	0.0	NEIN	NEIN
45	0.6	0.2	0.0	0.2	0.7	0.2	0.0	0.1	0.7	0.2	0.0	0.1	NEIN	NEIN
16	0.6	0.3	0.0	0.0	0.6	0.3	0.0	0.0	0.7	0.2	0.0	0.0	NEIN	NEIN
60	0.6	0.3	0.1	0.0	0.6	0.3	0.1	0.0	0.6	0.3	0.1	0.0	NEIN	NEIN
34	0.6	0.3	0.0	0.1	0.6	0.3	0.0	0.1	0.6	0.3	0.0	0.1	NEIN	NEIN
57	0.6	0.3	0.1	0.0	0.6	0.3	0.1	0.0	0.6	0.3	0.1	0.0	NEIN	NEIN
37	0.6	0.3	0.1	0.0	0.5	0.3	0.1	0.1	0.5	0.3	0.1	0.1	NEIN	NEIN
20	0.6	0.1	0.1	0.2	0.5	0.2	0.3	0.0	0.2	0.3	0.3	0.2	NEIN	NEIN
26	0.6	0.3	0.0	0.1	0.4	0.3	0.1	0.2	0.7	0.1	0.1	0.2	NEIN	NEIN
50	0.6	0.1	0.1	0.2	0.4	0.2	0.2	0.2	0.4	0.2	0.2	0.2	NEIN	NEIN
48	0.5	0.5	0.1	0.0	0.6	0.2	0.1	0.2	0.6	0.2	0.1	0.2	NEIN	NEIN
15	0.5	0.1	0.2	0.2	0.6	0.0	0.3	0.0	0.4	0.0	0.6	0.0	NEIN	JA
35	0.5	0.4	0.1	0.0	0.5	0.4	0.1	0.0	0.5	0.4	0.1	0.0	NEIN	NEIN
59	0.5	0.3	0.2	0.1	0.5	0.4	0.1	0.0	0.5	0.4	0.1	0.0	NEIN	NEIN
7	0.5	0.3	0.1	0.1	0.4	0.3	0.1	0.2	0.5	0.2	0.2	0.1	NEIN	NEIN
28	0.4	0.3	0.1	0.1	0.4	0.3	0.1	0.1	0.4	0.5	0.1	0.1	NEIN	NEIN
49	0.3	0.2	0.2	0.3	0.5	0.1	0.1	0.3	0.5	0.1	0.1	0.3	NEIN	NEIN
55	0.3	0.0	0.7	0.0	0.4	0.0	0.6	0.0	0.4	0.0	0.6	0.0	NEIN	NEIN
27	0.2	0.0	0.4	0.4	0.1	0.1	0.6	0.1	0.2	0.2	0.3	0.3	NEIN	NEIN
38	0.2	0.2	0.2	0.3	0.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	JA	NEIN
11	0.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0	NEIN	NEIN

11.3.3 Kaufhäufigkeiten auf individueller Ebene

In Tabelle 11.2 werden die Cue-Präferenzen anhand der prozentualen Häufigkeiten jeder Person ersichtlich. Dabei sieht man, dass der oben beobachtete Rückgang von Cue 1 in der zweiten Bedingung vor allem auf fünf Personen zurückzuführen ist (siehe zweitletzte Spalte: Vp 30, 33, 36, 38, 43). Bei weiteren 15 Versuchspersonen kann man nicht wirklich von einer Verhaltensänderung in der Cuesuche sprechen, aber sie reduzieren dennoch die Cuekäufe von Cue 1 und wählen häufiger einmal auch andere Cues. Diese Personen zeigen im zeitlichen Kontext eine breitere Streuung der Cuesuche als in der ersten Bedingung. Zehn Personen kaufen in der zweiten Bedingung den Cue 1 häufiger als in der ersten Bedingung. Ganze 30 Personen zeigen jedoch gar keine Anpassung an die dynamische Umwelt in ihren Häufigkeiten der Cue-Präferenzen. Die oben erwähnten Subgruppen zeigen sich somit dreigeteilt: 20 Personen, die ihre Kaufhäufigkeit von Cue 1 in der zweiten Bedingung reduzieren; 10 Personen, die diesen Cue häufiger kaufen und 30 Personen, die ihre Cue-Präferenz nicht anpassen.

Für eine Analyse des Einflusses der Cue-Stabilität wurden in Bedingung 2 und 3 die Häufigkeiten von Cue 1 / Cue 2 bzw. von Cue 3 / Cue 4 zusammengefasst. Dieses Vorgehen folgt der Annahme, dass die Person ihre Informationssuche nach der Validität ausrichtet, wenn sie Cue 1 oder 2 bevorzugt, und dass sie die Stabilität als Gütemass präferiert, wenn Cue 3 und 4 in der Suche überwiegen. Um eine Verhaltensänderung zwischen der zweiten und dritten Bedingungen zu detektieren, wurden die zusammengefassten Summen der prozentualen Häufigkeiten von Cue 1 und Cue 2 bzw. Cue 3 und Cue 4 miteinander verglichen. Unterschieden sich diese um mehr als .25 in ihrer Summe, dann wurde dieses Verhalten als Änderung kategorisiert (siehe letzte Spalte in Tabelle 11.2). Anhand dieser Auswertung lässt sich ein veränderndes Verhalten bei neun Personen definieren (Vp 3, 8, 10, 12, 15, 18, 22, 25, 29). Diese Personen richten ihre Informationssuche in der dritten Bedingung teilweise ausschliesslich, teilweise partiell nach der Stabilität der Cues. Zusätzlich erwerben weitere 30 Personen den Cue 3 auch ab und zu. Somit zeigen sich in der dritten Bedingung noch zwei Subgruppen: 39 Personen bevorzugten den stabilen Cue und 21 Personen den validen.

Bei den 39 Personen ist jedoch unklar, ob sie die Stabilitätsrate auch richtig anwendeten oder ob die bloße Angabe der Stabilität die Attraktivität dieses Cues erhöhte. Wir nehmen an, dass die Stabilität nicht verstanden wurde, wenn der stabilste Cue nach einer Konsultation gleich nochmals konsultiert wurde. Gegenläufig dazu gilt das Stabilitätskonzept als verstanden, wenn die Person den stabilen Cue konsultiert, ohne ihn in der Runde zuvor schon konsultiert zu haben oder ihn und gleich danach nochmals zu konsultieren. Die 39 Personen schauen den sehr stabilen Cue in der dritten Bedingung insgesamt dreimal häufiger an als in der zweiten Bedingung. Von den 653 mal, in denen er angeschaut wird, ist er 345 mal zweimal hintereinander konsultiert worden (hauptsächlich von 16 Personen) und 184 mal erfolgte dessen Konsultation ohne ihn vorher und nachher angeschaut zu haben (hauptsächlich von 14 Personen).

Tabelle 11.3: Häufigkeiten, mit denen die vier Cues als erstes, zweites, drittes etc. (Rangreihe 1 bis 8) angeschaut wurden; Angaben getrennt nach den drei Bedingungen bei den beiden Alternativen.

Bedingung 1								
Rang-reihe	Alternative X				Alternative Y			
	Cue 1	Cue 2	Cue 3	Cue 4	Cue 1	Cue 2	Cue 3	Cue 4
1	885	21	34	13	177	7	16	17
2	81	111	22	19	521	46	18	18
3	7	45	24	10	31	91	19	19
4	0	18	21	7	12	24	19	16
5	0	1	4	8	4	7	6	3
6	0	3	1	1	0	2	2	6
7	0	0	0	3	0	1	3	2
8	0	1	1	1	0	2	2	0

Bedingung 2								
Rang-reihe	Alternative X				Alternative Y			
	Cue 1	Cue 2	Cue 3	Cue 4	Cue 1	Cue 2	Cue 3	Cue 4
1	686	39	37	16	219	26	24	7
2	93	107	21	13	409	54	20	12
3	19	44	26	11	48	69	23	11
4	1	17	10	5	11	31	24	10
5	1	4	6	6	2	5	12	3
6	0	3	2	2	0	2	6	1
7	0	0	0	4	0	0	1	4
8	0	0	1	0	0	1	1	2

Bedingung 3								
Rang-reihe	Alternative X				Alternative Y			
	Cue 1	Cue 2	Cue 3	Cue 4	Cue 1	Cue 2	Cue 3	Cue 4
1	426	50	245	59	140	20	78	22
2	100	92	53	27	238	46	142	43
3	27	31	34	12	59	35	37	15
4	5	13	8	3	23	28	29	14
5	1	2	7	4	5	8	10	1
6	0	1	1	4	0	3	4	2
7	0	0	1	2	0	0	2	4
8	0	0	0	0	0	1	2	2

11.3.4 Suchregel

Die Häufigkeiten, mit denen die vier Cues bei den beiden Alternativen als erster, zweiter, dritter, vierter etc. angeschaut wurden, ist in Tabelle 11.3 dargestellt. Es ist zunächst auffällig, dass die Probanden eine starke Tendenz zeigen, in der Suche bei Alternative X zu beginnen. Sie steht im Information Board auf der linken Seite, so dass sich dieser Effekt durch die Gewohnheit erklärt, Dinge von links nach rechts zu lesen und zu bearbeiten. Interessant daran ist allerdings, dass der Grad der Asymmetrie von Bedingung 1 zur Bedingung 2 nachlässt: Während in Bedingung 1 das Verhältnis etwa 5:1 beträgt (885mal wird die Informationssuche mit Cue 1 von Alternative X gestartet, nur 177mal mit Cue 1 von Alternative Y), sinkt dieses Verhältnis durch die Einführung der Zeitachse auf etwa 3:1 (686mal Start mit Cue 1 der Alternative X gegenüber 219mal Start

mit Cue 1 der Alternative Y). In Bedingung 3 bleibt dieses Verhältnis konstant bei 3:1 (426mal Start mit Cue 1 der Alternative X gegenüber 140mal Start mit Cue 1 der Alternative Y, und, was als substantielle Zahl hinzutritt, 245mal Start mit Cue 3 der Alternative X gegenüber 78mal Start mit Cue 3 der Alternative Y). Dass die Einführung der Zeitachse eine Verringerung der Asymmetrie nach sich zieht, macht deutlich, dass sich zumindest ein Teil der Probanden in einem Teil der Durchgänge im Start der Informationssuche an den Informationen aus den vorhergehenden Handelstagen orientiert.

Tabelle 11.3 zeigt zudem, dass die Einführung der Cue-Stabilität in Bedingung 3 einen sehr deutlichen Einfluss auf den zuerst angeschauten Cue hat: Während Cue 3 (Validität .80, Stabilität mit .95 aber extrem hoch) in den Bedingungen 1 und 2 keine nennenswerte Rolle spielt, wird er in Bedingung 3 nicht nur gewählt, sondern sogar oft als erster oder zweiter Cue angeschaut (immerhin etwa halb so häufig wie der valideste Cue und etwa fünfmal häufiger als in Bedingung 1 oder 2).

Es wird aber auch deutlich, dass sich bereits in Bedingung 1 nicht alle Personen in allen Fällen rational verhalten. Es gibt ja keinen Grund, mit einem anderem als dem validesten Cue die Suche zu beginnen. Trotzdem finden sich immerhin 108 Fälle (bei einer Gesamtzahl von 1200 Fällen sind das 9%), in denen die Suche mit einem weniger validen Cue begonnen wurde. Diese Fälle verteilen sich auf 13 Personen, wovon besonders die Personen 11, 27, 38, 48, 49, 50, 51 und 55 in mehr als einem Drittel aller Fälle dieses Verhalten gehäuft zeigen. Bei einer Person (Vp 49) könnte möglich sein, dass sie von vornherein intendierte, mehrere Cues aufzudecken, so dass die Frage nach dem Start für sie keine Bedeutung spielte. Die sieben anderen stoppen jedoch häufiger nach einer Information als dass sie weitersuchen. Diese Fälle deuten darauf hin, dass diese Probanden möglicherweise das Validitätskonzept nicht richtig verstanden haben.

In Bedingung 2 erhöht sich die Zahl der Durchgänge, die nicht mit dem validesten Cue begonnen wurden, nur marginal (auf 149 oder 12,4%). Bei den individuellen Häufigkeiten in Tabelle 11.2 zeigte sich andeutungsweise eine breite Streuung der Informationssuche. Eine starke Berücksichtigung der vergangenen Informationen (die ja zumeist den zuvor aufgedeckten validesten Cue umfassten) hätte hier nun eine höhere Zuwachsrate erwarten lassen. Für den am ehesten dazu in Frage kommenden Cue 2 (von 28 auf 65 Fälle) ist diese auch zu beobachten, allerdings auf einem insgesamt zu niedrigen Häufigkeitsniveau (der Differenzbetrag von 37 Fällen ist gerade einmal 3,1% aller Durchgänge), als dass hier von einem substantiellen Strategiewechsel in der Stichprobe gesprochen werden kann. In der dritten Bedingung wird die Informationssuche insgesamt 151mal weder mit dem validesten noch dem stabilsten Cue gestartet.

Schliesslich ist zu Tabelle 11.3 zu bemerken, dass sich in den Häufigkeiten drei Stufen ergeben: Die eine nach dem ersten angeschauten Cue, die zweite nach dem zweiten und die dritte nach dem vierten angeschauten Cue. Vom ersten zum zweiten Cue fällt die Häufigkeit um ein Drittel, vom zweiten zum dritten Cue fällt die Häufigkeit etwa auf ein Drittel, genauso vom vierten bis zum fünften aufgedeckten Cue. Ein oder zwei Informationen scheinen also der Standardfall zu sein, in einer beträchtlichen Zahl der Fälle wird dann noch ein dritter oder vierter Cue ange-

schaut. Nach dem vierten ist dann aber die Informationssuche (von wenigen Einzelfällen abgesehen) beendet. Dieses Muster ist in den drei Bedingungen sehr stabil. Was man aus diesem Muster nicht herauslesen kann ist, ob es einzelne Aufgaben sind, die die Fortsetzung der Informationssuche über einen zweiten Cue hinaus nahelegen, oder ob sich Versuchspersonen in ihren Strategien interindividuell unterscheiden. Dies wird im Folgenden untersucht.

11.3.5 Stoppverhalten

In der Einleitung hatten wir drei verschiedene Stoppregeln definiert: Die *No New Reason Stopping Rule* (NNRStR) bezeichnet solche Fälle, in denen vor der jeweiligen Entscheidung keine neue Informationssuche stattfindet. Bei der *One Reason Stopping Rule* (ORStR) stoppt ein Proband die Informationssuche, sobald er eine diskriminierende Information gefunden hat, und bei der *More Reason Stopping Rule* (MRStR) setzt er die Informationssuche auch nach dem ersten diskriminierenden Cue noch fort. Da im vorliegenden Experiment die Informationen je Alternative einzeln gesucht wurden, und auch wegen der Asymmetrie der Nutzbarkeit positiver und negativer Vorhersagen, sind die möglichen resultierenden Fälle des Abbruchs einer Informationssuche vor der Auswertung der Daten diesen drei Stoppregeln zuzuordnen. Es gilt:

Keine Information in einem Durchgang gesucht (= NNRStR)

Stoppen nach dem ersten angeschauten Cue, gleich ob dieser positiv oder negativ ist (= ORStR)

Fortsetzen der Informationssuche, nachdem der erste angeschaute Cue zur einen Alternative negativ war, und Stoppen nach einem angeschauten Cue der anderen Alternative (gleich ob dieser positiv oder negativ ist) (= ORStR)

Suchen von mehr als einem Cue zur selben Alternative (= MRStR)

Suchen nach einem Cue zur anderen Alternative, nachdem zur einen Alternative bereits ein positiver Cue gefunden wurde (= MRStR)

Stoppverhalten auf der Ebene der Stichprobe. Als Grundmenge für die Auswertung auf Stichproben-Ebene gilt das Stoppverhalten von 1200 Entscheidungen (20 Runden und 60 Personen) pro Bedingung. In Bedingung 1 wird insgesamt 785mal gemäss der ORStR ($M_1 = 13.1$, $SD_1 = 5.9$) gestoppt, 385mal gemäss MRStR ($M_1 = 6.4$, $SD_1 = 6.2$) und 30mal ohne eine aktuelle Information ($M_1 = 0.5$, $SD_1 = 1.6$) entschieden. In der zweiten Bedingung wird die Informationssuche 672mal gemäss ORStR beendet ($M_2 = 11.2$, $SD_2 = 5.7$), 382mal gemäss MRStR ($M_2 = 6.4$, $SD_2 = 6.4$) und 146mal ohne eine aktuelle Information ($M_2 = 2.4$, $SD_2 = 4.0$) entschieden. In der dritten Bedingung wird 622mal anhand einer Information entschieden ($M_3 = 10.4$, $SD_3 = 5.6$), 418mal anhand mehreren Informationen ($M_3 = 7.0$, $SD_3 = 6.1$) und 160mal ohne jegliche Information ($M_3 = 2.3$, $SD_3 = 4.5$).

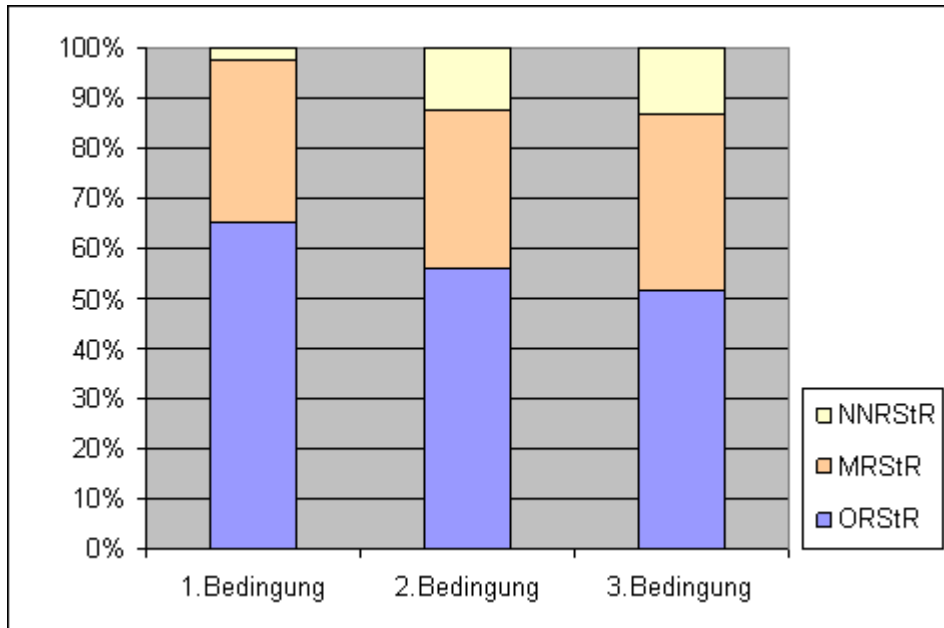


Abbildung 11.4: Kategorisierung des Stoppverhaltens in One Reason Stopping Rule (ORStR), More Reason Stopping Rule (MRStR) und No New Reason Stopping Rule (NNRStR).

Wie Abbildung 11.4 zusammenfassend zeigt, wurde die ORStR in allen drei Bedingungen in mehr als der Hälfte aller Entscheidungen angewendet. Es zeigt sich jedoch eine Verschiebung von der ersten zur zweiten Bedingung: Im zeitlichen Kontext wurde weniger häufig anhand einer Information entschieden als ohne Zeitkontext ($t = 3.52$; $p < .01$), etwa gleich häufig nach mehreren Informationen gesucht ($t = 0.11$; $p = .91$), aber signifikant häufiger ($t = 4.47$; $p < .01$) ohne jegliche Informationssuche entschieden. Die Verschiebung von der ersten zur zweiten Bedingung scheint also von ORStR zur NNRStR hin zu gehen. Es zeigen sich keine Veränderungen in den Häufigkeiten der angewendeten Stoppstrategien von der zweiten zur dritten Bedingung: ORStR mit $t = 1.65$; $p = 0.11$, MRStR mit $t = -1.34$; $p = .19$ und NNRStR mit $t = -0.78$; $p = .44$. Die anfangs auf Mittelwertebene beobachtete Abnahme der Kaufhäufigkeit von Bedingung 1 zu Bedingung 2 entsteht also vor allem dadurch, dass im zeitlichen Kontext häufiger ohne aktuelle Information entschieden wird. Ein Blick auf die individuellen Strategien soll klären, ob die gefunden Tendenz eine allgemeine ist oder durch einige Personen verursacht wird.

Stoppverhalten auf individueller Ebene. Die Zuteilung des Stoppverhaltens einer Versuchsperson zu einer bestimmten Strategie erfolgt gemäss Binomialtest ($\alpha = 0.05$): Ab mindestens 15 Treffern bei entweder ORStR oder MRStR gilt die jeweilige Stoppregel als dominant. Bei der NNRStR nehmen wir diese Strategie als dominant an, wenn mindestens 12mal ohne aktuelle Information entschieden wurde (Da bei den 20 Runden mindestens 4mal Informationen gekauft werden mussten und der Binomialtest deswegen von einem $n = 16$ auszugehen hat).

Tabelle 11.4: Anzahl Stoppverhalten, welche bei den 60 Vpn in den beiden Bedingungen gemäß One Reason Stopping Rule (O), More Reason Stopping Rule (M) und No New Reason Stopping Rule (N) klassifiziert werden. Die dominante Strategie ist fett markiert, sortiert sind die Daten nach abnehmender ORStR in der 1. Bedingung.

Vp-ID	1. Bedingung			2. Bedingung			3. Bedingung		
	O	M	N	O	M	N	O	M	N
53	20	0	0	20	0	0	20	0	0
14	20	0	0	20	0	0	20	0	0
13	20	0	0	20	0	0	20	0	0
32	20	0	0	20	0	0	19	1	0
10	20	0	0	19	1	0	20	0	0
52	20	0	0	19	1	0	11	9	0
58	20	0	0	19	0	1	7	13	0
8	20	0	0	17	3	0	0	6	14
39	20	0	0	16	0	4	18	0	2
46	20	0	0	14	1	5	16	1	3
33	20	0	0	12	1	7	10	1	9
55	20	0	0	10	0	10	8	0	12
41	20	0	0	4	0	16	4	0	16
23	19	1	0	20	0	0	18	2	0
22	19	1	0	19	1	0	16	4	0
25	19	0	1	16	0	4	14	1	5
31	19	0	1	14	0	6	17	0	3
36	19	0	1	13	6	1	17	3	0
17	19	1	0	12	8	0	11	9	0
3	18	0	2	13	0	7	13	0	7
6	17	3	0	17	3	0	15	5	0
54	16	4	0	17	3	0	11	9	0
56	16	4	0	15	1	4	15	2	3
40	16	4	0	12	8	0	14	4	2
24	16	3	1	10	1	9	10	1	9
5	15	0	5	16	0	4	13	1	6
27	15	5	0	16	3	1	10	10	0
4	15	5	0	9	3	8	13	2	5
43	15	0	5	6	0	14	5	0	15
30	15	4	1	4	16	0	2	17	1
18	14	6	0	19	1	0	16	4	0
12	14	4	2	5	12	3	2	14	4
37	13	7	0	9	11	0	9	11	0
48	12	8	0	13	7	0	11	9	0
51	12	8	0	10	6	4	10	9	1
47	11	9	0	13	7	0	13	7	0
50	11	9	0	13	7	0	8	10	2
29	11	9	0	10	3	7	7	3	10
38	10	10	0	17	0	3	17	2	1
60	10	10	0	10	10	0	11	9	0
2	10	10	0	8	12	0	12	8	0
11	10	0	10	5	0	15	4	0	16
45	9	10	1	11	8	1	12	1	7
15	8	12	0	7	8	5	7	9	4
42	7	13	0	12	3	5	6	12	2
16	7	13	0	8	12	0	11	9	0
9	7	13	0	7	13	0	12	7	1
44	7	13	0	6	14	0	7	13	0
21	6	14	0	10	10	0	10	10	0
1	6	14	0	7	13	0	9	11	0
34	6	14	0	6	14	0	5	15	0
19	6	14	0	2	18	0	0	20	0
20	5	15	0	5	15	0	6	14	0
49	5	15	0	3	17	0	4	16	0
57	4	16	0	4	16	0	0	20	0
59	4	16	0	2	18	0	0	20	0
26	4	16	0	1	19	0	8	12	0
28	3	17	0	6	12	2	10	10	0
7	3	17	0	2	18	0	8	12	0
35	2	18	0	2	18	0	0	20	0

Wie Tabelle 11.4 zeigt, ist in Bedingung 1 die häufigste dominante Stopppstrategie die ORStR (30 Versuchspersonen, also genau 50% der Stichprobe). Nur acht Probanden setzen signifikant am häufigsten die MRStR ein.

Dieser Befund ist konsistent mit früheren, vergleichbaren Experimenten, jedoch ist die Asymmetrie zugunsten der ORStR wesentlich deutlicher ausgeprägt. Wenn man jedoch die hohen Validitäten berücksichtigt und annimmt, dass damit die Schwelle der Urteilssicherheit bei praktisch allen Personen überschritten sein dürfte, dann wird diese Asymmetrie zugunsten des *One Reason Decision Making* erklärlich.

Interessant ist im Kontext der vorliegenden Untersuchung aber nun, wie die Personen auf die Einführung der Zeitachse (Bedingung 2) reagieren. Insgesamt gibt es von den 60 Probanden drei, welche sich radikal auf das Minimum an notwendiger Information umstellen und eine praktisch konsequente NNRStR zeigen (mindestens viermal mussten sie ja zumindest einen Cue kaufen, um nicht zu raten). 15 weitere Personen vertrauen zumindest in einigen Fällen auf die vergangene Information (Minimum 4x NNRStR). Nur eine von ihnen (Vp 5) hatte bereits in Bedingung 1 mehrmals ein solches Verhalten gezeigt, die anderen 14 scheinen dies aufgrund der nun verfügbaren Information aus der Vergangenheit zu tun. Damit reagieren insgesamt 17 der 60 Versuchspersonen mit dem Einsetzen der NNRStR auf die dynamische Umwelt.

13 dieser 17 Personen gehören zu der Hälfte Probanden, die in der ersten Bedingung eine ORStR gezeigt hatten. Von den acht Personen, die in Bedingung 1 zur MRStR gerechnet wurden, wendet keine einzige mindestens 4x die NNRStR in Bedingung 2 an. Selbst wenn man hier das Kriterium lockert und weitere acht Personen einberechnet, die in Bedingung 1 zumindest 14- oder 13mal gemäss MRStR gestoppt hatten, findet sich nur eine einzige Person, die mehrfach die NNRStR in Bedingung 2 einsetzt (Vp 42). (Diese Person erweist sich auch noch als Wechsler hin zu einer ORStR in Bedingung 2). Die NNRStR in Bedingung 2, die mit der Zeitachse auch Informationen aus der Vergangenheit berücksichtigt, scheint also vor allem für *One Reason Decision Maker* eine Option zu sein. *More Reason Decision Maker* hingegen nutzen die Information aus der Vergangenheit nicht. Sie verwenden die Information aus der Vergangenheit auch nicht, um substantiell häufiger auf das sparsamere ORDM auszuweichen: Während die 15 Personen der „erweiterten“ MRDM-Gruppe in Bedingung 1 insgesamt 75mal die ORStR angewendet hatten, war dies in Bedingung 2 mit 71mal sogar noch etwas seltener der Fall.

Wie intraindividuell stabil diese Strategien zwischen den beiden Umwelten sind, zeigt sich daran, dass es bei den 30 *One Reason Decision Makers* zur Bedingung 2 hin nur einen echten „Strategiewechslers“ gibt (Vp 30 von ORStR zu MRStR), dafür aber 17, die ihrer Strategie treu bleiben. Zehn Personen adaptieren ein wenig an die dynamische Umwelt und basieren ihre Entscheidung nicht mehr ausschliesslich auf eine Information, zwei Personen betreiben konsequent NNRStR. Von den acht konsequenten *More Reason Decision Makers* in der ersten Bedingung, verändern sieben ihre Strategie nicht signifikant und eine Person verwendet beide in der zweiten Bedingung. Die restlichen 22 Personen verwenden in beiden Bedingungen beide Stopppstrategien

uneinheitlich, wobei vor allem eine Person noch auffällt, die von einer gemischten Strategie zu einem eindeutigen ORStR in der zweiten Bedingung wechselt (Vp 38).

Über alle Auswertungen gesehen lässt sich ein klarer Zusammenhang zwischen den Subgruppen der Stoppregeln (*ORDM* vs. *MRDM*) und den Subgruppen der Informationssuche finden (vergleiche dazu die Tabellen 11.2 und 11.4). Von den fünf Personen, welche den Cue 1 in der zweiten Bedingung weniger häufig kaufen, sind vier *One Reason Decision Makers* in der ersten Bedingung (und einer verwendet beide Stoppstrategien gleich häufig). Von den 9 Personen, die Cue 3 in der dritten Bedingung hauptsächlich beachten, sind fünf Personen *One Reason Decision Makers* in Bedingung 1. (Die restlichen vier lassen sich in der ersten Bedingung keiner der Subgruppen zuordnen). Die Gruppe der *More Reason Decision Makers* fehlt vollständig bei denjenigen Personen, die durch die Zeitachse ihr Informationssuchverhalten massgeblich ändern.

11.4 Diskussion

Interexperimentell stabiles Informationssuchverhalten. Die statische Vergleichsbedingung des Experiments (Bedingung 1) repliziert im Wesentlichen die Befunde, die man aus vergleichbaren Experimenten zur Informationssuche im Zweialternativenfall kennt: Die Personen befragen die Cues in absteigender Validitätsfolge, aber nicht immer. (In unserem Fall haben wir 9% „Ausreisser“ von dieser Suchregel). Eine Subgruppe stoppt entweder nach dem ersten positiven Cue bzw. nach den ersten Hinweisen, dass beide Optionsscheine sich negativ entwickeln werden (*One Reason Stopping Rule*), eine andere Subgruppe sucht trotz vorliegender Information weiter (*More Reason Stopping Rule*). Ferner gibt es, wie in diesen Experimenten üblich, auch diesmal eine ganze Reihe von Personen, deren Stoppverhalten bezüglich der theoretisch postulierten Strategien so uneinheitlich ist, dass man sie besser nicht klassifizieren sollte. In Such- und Stoppregel wiederholen sich also die bekannten Resultate. Dies ist nur deswegen eine Bemerkung wert, als im vorliegenden Experiment ja die Cuesuche nicht wie bei vorausgehenden Experimenten jeweils einen Hinweis für alle Alternativen gab, sondern immer nur für eine. Diese Veränderung hätte die Möglichkeit ergeben, bei den Probanden ein ganz anderes, z.B. alternativenweises Such- und Stoppverhalten zu provozieren.

Auswirkungen der Zeitachse. Am Ende der Einleitung wurden drei Hypothesen formuliert, wie sich die Einführung der Zeitachse und die Einführung der damit verbundenen Stabilitätsmasse auswirken sollten, wenn die Probanden diese neue Situation adäquat berücksichtigen.

Alle drei Hypothesen konnten durch die Resultate bestätigt werden: Der Gesamtumfang gesuchter Information sinkt in Bedingung 2 markant ab (Hypothese 1), was wir als ein Indiz deuten, dass die Probanden nun auch die Informationen aus der Vergangenheit einbeziehen und deswegen weniger neue Information benötigen. In der Tat lässt sich dieser Effekt in Bedingung 2 zu einem guten Teil auf Fälle zurückführen, in denen *One Reason Decision Makers* in einigen Durchgängen aufgrund der Informationen aus der Vergangenheit ein dermassen gutes Bild von der aktuellen Situation zu haben glauben, dass sie ganz auf Informationssuche verzichten (Hypothese 2). Und mit der Einführung von Wissen über die Cue-Stabilität in Bedingung 3 verlagern

die Versuchspersonen – oder zumindest ein Teil von ihnen – die Informationssuche in einer Reihe von Fällen auf die weniger validen, dafür aber stabileren Cues, und zwar sowohl in der Häufigkeit (Hypothese 3a) als auch in der Suchreihenfolge (Hypothese 3b).

Als wichtigstes Resultat zeigte sich, dass sich bei der Einführung der Zeitdimension neben der Informationssuche vor allem die Stoppregel für die Informationssuche ändert: Es werden vermehrt auf der Basis von vergangenen Informationen Entscheidungen gefällt. Es gibt zahlreiche Personen, die nicht immer eine aktuelle Information benötigen und häufiger auch mal keine Information kaufen. Jedoch wirkt sich – und das ist die wichtigste Einschränkung, die sich aus den Analysen auf individueller Ebene ergibt – die vergangene Information nur für ein knappes ein Drittel der Probanden als bedeutsam aus. Für die Mehrheit der Stichprobe liesse sich interpretieren, dass sie a) bei ihren Entscheidungen lieber auf Nummer sicher gehen und sich ungern auf vergangene (und damit unsicherere) Informationen verlassen oder b) ihre Informationssuchstrategie nicht anpassen, weil sie die Veränderungen der statistischen Umwelt nicht richtig erfassen.

Dass sich unter denjenigen Personen, die ihr Suchverhalten substantiell auf die spezifische statistische Umwelt der Zeitachse einstellen, keine einzige Person aus der Gruppe der *More Reason Decision Makers* in Bedingung 1 befindet, interpretieren wir durch das Modell eines erhöhten Sicherheitsbedürfnisses, welches bei von Personen sowohl zu MRDM als auch zu einer gewissen Ablehnung von über die Zeit doch nicht mehr so validen Cue-Informationen. Das Nutzen von Informationen aus der Vergangenheit oder für die Zukunft hingegen wäre dann für Personen attraktiv, welche eher an einer Minimierung des Suchaufwandes als an einer Maximierung der Entscheidungssicherheit interessiert sind.

Revival der Minimalist-Heuristik. Wir möchten an dieser Stelle gerne noch einmal auf den Befund zurückkommen, dass sich einige der *One Reason Decision Makers* aus Bedingung 1 in Bedingung 2 derart stark auf die Informationen aus der Vergangenheit beziehen, dass sie hinsichtlich der Cuegüte als „Minimalisten“ klassifiziert werden müssen (zumindest sich dies die beiden Vpn 41 und 43, aber man auch noch diejenigen sieben Personen ganz oder teilweise hinzurechnen, bei denen aufgrund zu vieler Entscheidungen ganz ohne Informationen eine Zuordnung zur ORStR nicht mehr signifikant möglich war). Diese zwei bis neun Personen sind nicht etwa wegen der Grösse dieser Gruppe für eine Diskussion interessant, sondern deswegen, weil sie ja nicht nur gegen Stoppregel, sondern auch gegen die Suchregel von *Take The Best* verstossen. TTB fordert, Cues in der Reihenfolge ihrer Validitäten zu befragen. Diese Reihenfolge wurde, wie wir bei den Resultaten gesehen haben, auch in Bedingung 1 in der Regel eingehalten. (Eigentlich gab es dort auch keine vernünftige Alternative dazu, da das Experiment ja die Diskriminationsrate auf 100% gesetzt hatte). In Bedingung 2 finden wir nun Personen, die systematisch dagegen verstossen. Wenn sie dies zu ihrer Hauptstrategie machen (so wie die Probanden 41 und 43), dann ist eine Klassifikation als „Minimalisten“ zutreffend. Die Minimalist-Heuristik unterscheidet sich von TTB darin, dass die Cues zufällig und nicht nach ihrer Validität gesucht werden. Bislang lief diese Strategie in der Theorie zu Urteilsheuristiken immer unter „ferner liefern“ und wurde eher als einer der weiteren möglichen Strategien aufgelistet (z.B. Czerlinski, Gigerenzer, & Goldstein, 1999; Katsikopoulos & Martignon, 2006; Todd & Dieckmann, 2005). Jedoch gab

es keine experimentellen Settings, in denen eine Anwendung dieser Strategie sinnvoll gewesen wäre. Mit der Einführung der Zeitachse liegt nun ein solches Setting vor: Wenn man bereit ist, sich ganz auf Informationen aus der Vergangenheit zu verlassen, dann verliert man an Validität, gewinnt aber an Investitionschancen (denn man gibt ja weniger Geld für Informationssuche aus). Für risikosuchende Personen wird Minimalist also unter reinem Kosten-Nutzen-Aspekt in einer zeitlich dynamischen Welt genau so zu einer Option wie für extrem Sicherheitsbewusste, die ihr Geld lieber auf das Gewinnkonto transferieren als für Informationssuche einzusetzen (letzteres vor allem dann, wenn die Informationen aus der Vergangenheit negativ waren).

Dass Minimalist als eine Strategie in zeitlich dynamischen Welten an Bedeutung gewinnt, zeigt sich noch deutlicher in Bedingung 3 des Experiments. Gegeben einmal der Annahme, dass die Versuchspersonen die weniger validen, dafür aber sehr stabilen Cues wirklich mit Blick auf die Zukunft konsultieren, dann tun sie dies unter Preisgabe der TTB-Suchregel, welche streng absteigende Validität fordert. Ohne dass wir auf der Basis eines einzigen Experimentes bereits einen quantitativen Effekt bemühen möchten, ist die Zahl der Fälle, in denen zunächst Cue 3 oder Cue 4 konsultiert wurden, doch so erheblich, dass dies auf eine Besonderheit der präsentierten Umwelt hindeutet: Die Personen setzen nicht mehr nur auf Validität. Bislang eher ein Mauerblümchen, so könnte Minimalist also im Rahmen zeitlich dynamischer Umwelten an Bedeutung gewinnen.

Es sei angemerkt, dass diese Diskussion von Minimalist in Abgrenzung zu TTB natürlich nicht gegen TTB an sich spricht, denn diese Urteilsheuristik ist ja formuliert als optimiertes Vorgehen in Fällen, in denen Informationssuche keine Ressourcen kostet, in denen aber auf eine Begrenzung der Verarbeitungsressourcen (z.B. Arbeitsgedächtnis oder computationale Fähigkeiten) Rücksicht genommen werden muss. Wir betreten mit den zeitlich dynamischen Umwelten bei gleichzeitiger Berücksichtigung von Informationssuchkosten ein ganz anderes Feld, in dem andere Gesetzmässigkeiten gelten könnten. Und es zeigt sich ja in den Resultaten, dass auch hier Validität immer noch die hauptsächliche Triebfeder der Informationssuche bleibt.

Schlussbemerkung. Insgesamt gesehen wäre es vermessen, den Resultaten eines ersten Experimentes in ihrem quantitativen Gefüge zueinander eine übermässig grosse Bedeutung beizumessen. Mit den aufgestellten Hypothesen ist vielmehr das Feld definiert, welches bei der Berücksichtigung zeitlich dynamischer Umwelten von der Entscheidungsforschung zu bearbeiten ist. Und dass sich die Hypothesen bestätigen liessen, ist ein klares Indiz dafür, dass Menschen sensitiv auf solche zeitlich dynamische Umwelten reagieren: Zumindest einige suchen nicht nur gegenwartsorientiert Informationen, sondern berücksichtigen auch die Vergangenheit und teilweise sogar die Zukunft. Es wäre jedoch überzogen, die doch je spezifischen Restriktionen eines einzelnen experimentellen Designs unberücksichtigt zu lassen und die Resultate deswegen auf Informationssuche in zeitlich dynamischen Systemen im Allgemeinen zu erweitern. Unterstrichen haben sollte die vorgelegte Studie aber, dass hier eine interessante Forschungsaufgabe besteht, um die Entscheidungsforschung in Experiment und Theoriebildung näher an die realen Bedingungen menschlichen Entscheidens (in der Zeit nämlich) heranzuführen.

11.5 Literatur

- Betsch, T., Haberstroh, S., Glöckner, A., Haar, T., & Fiedler, K. (2001). The effects of routine strength on adaption an information search in recurrent decision making. *Organizational Behavior and Human Decision Processes*, 84(1), 23-53.
- Bröder, A. (2000a). Assessing the empirical validity of the "Take The Best"-heuristic as a model of human probabilistic inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 1332-1346.
- Bröder, A. (2000b). *"Take the best - Ignore the rest" Wann entscheiden Menschen begrenzt rational?* Lengerich: Pabst Science Publishers.
- Bröder, A., & Schiffer, S. (2006). Adaptive flexibility and maladaptive routines in selecting fast and frugal decision strategies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32, 904-918.
- Bussemeyer, J. R., & Townsend, J. T. (1993). Decision Field Theory: A Dynamic-Cognitive Approach to Decision Making in an Uncertain Environment. *Psychological Review*, 100(3), 432-459.
- Christen, S., Zurbruggen, S., Hausmann, D., & Läge, D. (2007). *Trinkwasser- und Börsenszenario: Aufbau, Ablauf und technische Spezifikation einer Experimentalumgebung zur entscheidungsvorbereitenden Informationssuche in zeitlichen Verläufen*. AKZ-Forschungsbericht 43, Zürich: Angewandte Kognitionspsychologie.
- Czerlinski, J., Gigerenzer, G., & Goldstein, D. (1999). How good are simple Heuristics? In G. Gigerenzer, P. Todd & A. R. Group (Eds.), *Simple Heuristics that make us smart*. New York, Oxford: Oxford University Press.
- Gigerenzer, G., & Goldstein, D. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, 103, 650-669.
- Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review*, 98, 506-528.
- Hausmann, D., Christen, S., & Läge, D. (2006). Urteilssicherheit um jeden Preis? Verletzung der Stopppregel in Abhängigkeit von der Höhe der Informationssuchkosten. In B. Gula, R. Alexandrowicz, S. Strauss, E. Brunner, B. Jenull-Schiefer & O. Vitouch (Eds.), *Perspektiven psychologischer Forschung in Österreich. Proceedings zur 7. Wissenschaftlichen Tagung der Österreichischen Gesellschaft für Psychologie* (pp. 45-51). Lengerich: Pabst.
- Hausmann, D., & Läge, D. (2005). How the desired level of confidence influences the stopping rule of information search. In K. Opwis & I. Penner (Eds.), *Proceedings of KogWis05. The German Cognitive Science Conference 2005* (pp. 71-76). Basel: Schwabe.
- Katsikopoulos, K. V., & Martignon, L. (2006). Naïve heuristics for paired comparisons: Some results on their relative accuracy. *Journal of Mathematical Psychology* 50(5).
- Kerstholt, J. H., & Raaijmakers, J. G. W. (1997). Decision making in dynamic task environments. In W. R. Crozier & O. Svenson (Eds.), *Decision Making: Cognitive Models and Explanations* (pp. 205-217). London: Routledge.
- Läge, D., & Hausmann, D. (2007). *Der Einfluss der Informationssuche auf die Entscheidung*. AKZ-Forschungsbericht 04, Zürich: Angewandte Kognitionspsychologie.
- Läge, D., Hausmann, D., & Christen, S. (in prep). *Information search strategies are learnable: validity, discrimination rate, usability, and success*.
- Newell, B., Rakow, T., Weston, N., & Shanks, D. (2004). Search strategies in decision making: The success of 'success'. *Journal of Behavioral Decision Making*, 17, 117-137.
- Newell, B., & Shanks, D. (2003). Take the best or look at the rest? Factors influencing one-reason decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 53-65.

- Newell, B., Weston, N., & Shanks, D. (2003). Empirical tests of a fast and frugal heuristic : not everyone "takes-the-best". *Organizational Behavior and Human Decision Processes*, 91, 82-96.
- Payne, J. W. (1976). Task complexity and contingent processing in decision making: An information search and protocol analysis. *Organizational Behavior and Human Performance*, 16, 366-387.
- Rakow, T., Newell, B., Fayers, K., & Hersby, M. (2005). Evaluating Three Criteria for Establishing Cue-Search Hierarchies in Inferential Judgment. *Journal of Experimental Psychology / Learning, Memory & Cognition*, 31(5), 1088-1104.
- Todd, P. M., & Dieckmann, A. (2005). Heuristics for ordering cue search in decision making. In L. K. Saul, Y. Weiss & L. Bottou (Eds.), *Advances in neural information processing systems* (Vol. 17). Cambridge, MA [et al.]: MIT Press.
- Zurbriggen, S., Christen, S., Hausmann, D., & Läge, D. (2008). *Adaptivität der Informationssuche in statischen und in dynamischen Umwelten*. AKZ-Forschungsbericht 46, Zürich: Angewandte Kognitionspsychologie.
- Zurbriggen, S., Hausmann, D., Christen, S., & Läge, D. (2007). *Die zeitliche Dynamik von Umwelten: Heuristisches Wahlverhalten am Beispiel eines Trinkwasser-Szenarios mit und ohne Zeitachse*. AKZ-Forschungsbericht 41, Zürich: Angewandte Kognitionspsychologie.
- Zurbriggen, S., & Läge, D. (2007). *Ressource Zeit statt Geld: Ein Vergleich zweier Szenarien zur Entscheidungsforschung*. AKZ-Forschungsbericht 45, Zürich: Angewandte Kognitionspsychologie.